Computing the vulnerability of time-evolving networks to infections

Eugenio Valdano

**THESE DE DOCTORAT DE
L'UNIVERSITE PIERRE ET MARIE CURIE**

Spécialité Biomathématiques

École Doctorale Pierre Louis de Santé Publique à Paris :
Épidémiologie et Sciences de l'Information Biomédicale

Présentée par

M. Eugenio VALDANO

Pour obtenir le grade de

DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Analyse quantitative de la vulnérabilité
des réseaux temporels aux maladies infectieuses

à soutenir le 13 octobre 2015 devant le jury composé de :

M. le Docteur Alain BARRAT (CNRS Marseille)	Examinateur
Mlle le Docteur Vittoria COLIZZA (Université Pierre et Marie Curie)	Co-directrice de thèse
M. le Docteur Philipp HÖVEL (Technische Universität Berlin)	Rapporteur
Mme le Professeur Clémence MAGNIEN (Université Pierre et Marie Curie)	Examinateur
M. le Professeur Guy THOMAS (Université Pierre et Marie Curie)	Directeur de thèse
Mme le Docteur Elisabeta VERGU (INRA)	Rapporteur
M le Professeur Alessandro VESPIGNANI (Northeastern University Boston)	Examinateur

Résumé de la thèse

Introduction

Les modèles mathématiques représentent une contribution importante dans la lutte contre les maladies infectieuses, comme ils représentent un outil pour définir des stratégies pour réduire le nombre de cas, utiliser de façon optimale des ressources limitées, faire face rapidement aux émergences, et mettre en place des mesures d'endiguement efficaces. Depuis leur introduction, ces modèles, qu'on appelle *modèles à compartiments*, sont basés sur la supposition que l'évolution et la transmission de la maladie puissent être traduites en un ensemble de règles simples et qui peuvent être adaptées à des agents pathogènes caractérisés par différents pathophysiologie, développement clinique, et étiologie (virus, bactéries, etc.). L'épidémie est donc interprétée comme un comportement émergent de l'interaction parmi les hôtes [1–5]. Néanmoins, dans le passé, l'utilisation de modèles mathématiques en santé publique a été limitée par le manque de données sur les interactions humaines. Récemment, la situation a radicalement changé, grâce à l'avènement de la *science des données*, grâce à laquelle il est devenu possible d'enregistrer de façon très précise les contacts et les transports responsables de la propagation des maladies. Actuellement, on a des bases de données qui décrivent l'interaction entre personnes, dans différents contextes et à différentes échelles : des contacts sociaux à la mobilité, jusqu'aux contacts sexuels [6–34]. En outre, les activités humaines ne sont pas le seul objet de recherche : il est maintenant possible de suivre les mouvements d'animaux de ferme [35–40], afin d'étudier la propagation des nombreuses maladies qui menacent la santé des animaux, l'économie, et finalement même la santé humaine.

Épidémies et réseaux

Afin de gérer ces structures des contacts complexes, de nouveaux outils théoriques ont été adapté et développé, qui considèrent la propagation de maladies comme un processus dynamique sur réseaux [41–48]. Les réseaux sont en fait devenus un outil essentiel qui, en représentant les populations en tant de nœuds (hôtes potentiels de l'agent pathogène) et des liens (interactions entre hôtes), permettent d'appliquer au contexte des maladies infectieuses le formalisme de la physique statistique, la théorie des graphes, et des systèmes dynamiques.

La dynamique épidémique sur réseau suppose que chaque nœud peut se trouver dans une des stades de la maladie prédites par le modèle à compartiments de référence, et que le processus de l'infection se déroule par les liens : chaque nœud a une probabilité d'infecter (transmissibilité) un nœud susceptible auquel il est connecté.

Un aspect clé dans la modélisation épidémique concerne l'émergence du régime épidémique. Quand un agent pathogène est introduit dans une population susceptible, deux scénarios sont possibles : soit le nombre d'infectés va vite à zéro, en conduisant rapidement à l'extinction, soit il augmente de façon exponentielle, ce qui entraîne une flambée

épidémique. Être capable de prédire lequel des deux aura lieu est fondamental pour estimer la vulnérabilité de la population attaquée par un agent pathogène spécifique, et ensuite pour évaluer les stratégies les plus efficaces d'immunisation. Ce concept de vulnérabilité de la population est formalisé en termes de *seuil épidémique* [4, 41], qui est la valeur critique de la transmissibilité qui distingue extinction et régime épidémique. L'épidémiologie computationnelle a fourni des outils pour calculer analytiquement la valeur du seuil épidémique pour de nombreux modèles épidémiques, et pour de différents types de structures de contact. En particulier, grâce à l'intégration de la théorie des réseaux, il a été possible d'évaluer l'impact des différentes caractéristiques topologiques sur la vulnérabilité, comme la présence de nœuds avec un degré (nombre de contacts) exceptionnellement élevé, qu'on appelle *hubs*. Il a été montré, en effet, que leur présence augmente considérablement la vulnérabilité du réseau, en réduisant le seuil épidémique. Pour montrer cela analytiquement, on va supposer un modèle susceptible-infecté-susceptible (SIS), dans lequel les nœuds infectés transmettent la maladie aux voisins susceptibles avec une probabilité λ, et qui guérissent avec une probabilité μ. Ce modèle est clairement une simplification du développement clinique réel d'une maladie ; il représente néanmoins un modèle utile et applicable à des maladies qui admettent l'existence d'un état endémique, comme un grand nombre d'infections bactériennes. On suppose en outre que $P(k)$ soit la distribution des degrés du réseau. Le seuil épidémique (λ_c) peut donc s'exprimer en termes des moments de cette distribution [41, 49, 50] : $\lambda_c = \frac{\mu\langle k\rangle}{\langle k^2\rangle}$; vu que la présence de hubs induit de fortes variations de degré, le second moment de la distribution est très grand. Cela signifie que le seuil est très bas, et diminue de plus en plus avec l'augmentation de la taille de la population. Le valeur du seuil dans ce type d'approche découle des propriétés statistiques des classes de réseaux (*annealed*) ; il est également possible de prendre en compte la topologie explicite du réseau de contacts à travers le formalisme de la matrice d'adjacence (*quenched*). En supposant un réseau de N nœuds, indicés $i, j = 1, \cdots, N$, on définit la *matrice d'adjacence* [51] A, de taille $N \times N$, et d'éléments $A_{ij} = 1$ si i est connecté à j, et $A_{ij} = 0$ autrement. Le seuil épidémique peut ansi être calculé en termes du spectre de la matrice d'adjacence [52, 53] :

$$\lambda_c = \frac{\mu}{\rho[A]}, \tag{1}$$

où $\rho[A]$ est le rayon spectral de la matrice d'adjacence, soit le maximum parmi les valeurs absolues de ses valeurs propres.

Seuil épidémique sur des réseaux temporels

Les résultats décrits ci-dessus ont été développées en supposant que les maladies se propagent sur des réseaux qui ne varient pas dans le temps, ou qu'ils le fassent à des échelles de temps très différentes de celles de la diffusion de l'agent pathogène. Dans de différents contextes, il a cependant été observé que les contacts et l'épidémie évoluent sur des temps caractéristiques comparables ; il a également été montré que l'interaction entre l'évolution du réseau et le processus de l'infection influence sensiblement le résultat du processus

épidémique [13, 54–59], les conditions de déclenchement, de persistance et d'endémicité. Par conséquent, une méthodologie qui aspire à une évaluation quantitative du risque doit tenir compte de l'évolution temporelle de contacts. Le formalisme des *réseaux temporels* (ou *réseaux dynamiques*) [60–62], où le réseau lui-même devient un processus dynamique qui détermine l'évolution des liens, peut être utilisé efficacement pour représenter ce type de structure ; cependant, il a été noté que les propriétés émergentes de ces nouveaux systèmes sont conceptuellement et phénoménologiquement nouvelles, et ne peuvent généralement pas être considérées comme de simples extensions de ce qui est connu pour les réseaux statiques. En particulier, l'étude du seuil épidémique a été limitée à des approches numériques ou des approches analytiques pour des cas spécifiques [59, 63–67]. Dans cette thèse, nous proposons une nouvelle méthode, publiée dans [68, 69], pour calculer analytiquement le seuil épidémique pour un réseau temporel générique. Le concept fondamental derrière cette méthode est de représenter le réseau en termes d'un objet statique : le réseau multi-couches (ou multi-layer) [70, 71]. Dans ce type d'objet les nœuds sont divisées en plusieurs couches, et les liens qui joignent des nœuds de la même couche ou de couches différentes sont interprétées de manière différente. On propose une façon innovatrice pour représenter le réseau temporel, où chaque pas de temps est représenté comme une couche qui contient une copie de tous les nœuds du réseau, et les liens du réseau temporel se traduisent en termes de liens entre les couches du réseau multi-couche. Les règles de construction du réseau multi-couche sont illustrées dans la Figure R1. La structure spécifique de cette projection permet de prendre en compte aussi la dynamique de l'infection. Le résultat est un objet qui représente à la fois le processus dynamique de l'évolution structurelle des contacts, et le processus de propagation de l'épidémie. Cet aspect nous a permis de résoudre analytiquement le calcul du seuil épidémique ; en particulier, on montre qu'il peut être dérivé à partir de ce que nous appelons *propagateur d'infection* [69]. En supposant un réseau temporel de T pas de temps, chacun avec matrice une d'adjacence A_t, le propagateur d'infection a la forme suivante :

$$P = \prod_{t=1}^{T} \left(1 - \mu + \lambda A_t\right). \tag{2}$$

On montre que le seuil épidémique est la valeur minimale de transmissibilité λ_c pour laquelle $\rho[P(\lambda_c, \mu)] = 1$ [72]. Le propagateur d'infection représente en fait le potentiel de propagation, puisque l'élément P_{ij} est égale à la probabilité que le nœud i, infectieux au temps $t = 1$, donne lieu à un nœud j qui soit infectieux au temps $t = T$, dans l'approximation *quenched mean field* [52, 53, 73] avec laquelle on l'a calculé.

En [68] on n'a introduit cette méthode que pour le modèle de diffusion SIS ; dans [69] on montre que la forme du propagateur d'infection décrite dans l'Equation 2 est valable même pour des maladies qui confèrent l'immunité après la guérison, soit temporaire soit permanente. La précision de notre mesure du seuil peut être estimée avec des simulations numériques de l'infection, en mesurant la fraction de nœuds infectés en fonction de la transmissibilité, lorsque la maladie atteint l'état endémique : la Figure R2**A** montre un exemple de cette approche.

Étant donné que les bases des données disponibles qui caractérisent les réseaux de
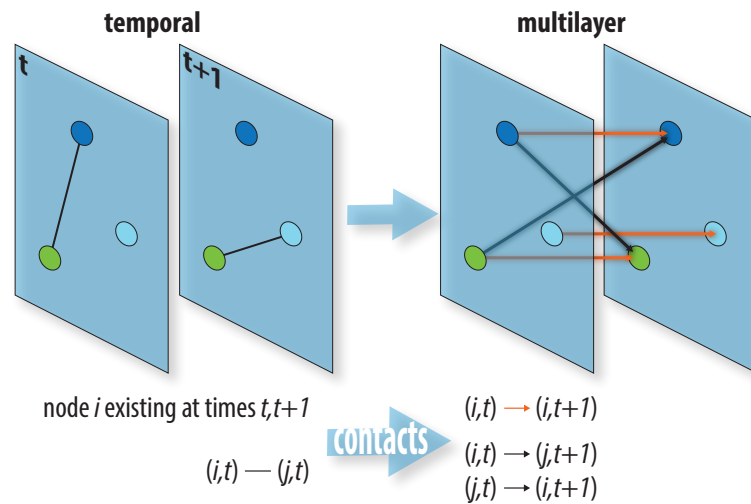
FIGURE R1: **Représentation multi-couche du réseau temporel.** Ici sont décrites les règles pour construire notre représentation multi-couches du réseau temporel, en utilisant seulement trois nœuds et deux pas de temps. Les liens qui joignent des représentations du même nœud sont pondérées avec $1 - \mu$, afin de représenter le processus de guérison – plus précisément de non-guérison – en termes de transmission de la maladie à soi-même dans le futur ; par contre, les liens qui représentent les contacts du réseau temporel sont pondérés avec la transmissibilité λ.
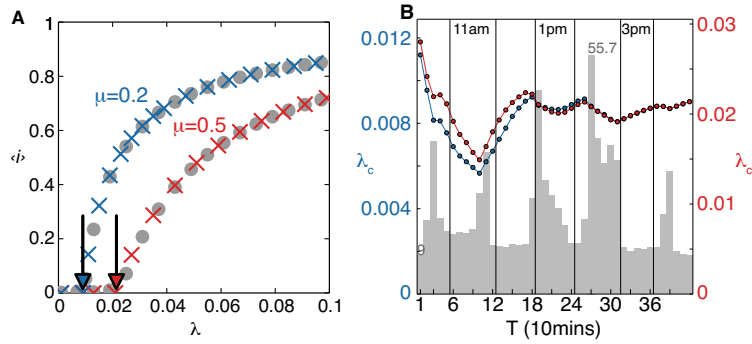
FIGURE R2: **Validation du calcul du seuil, et influence du temps total du réseau sur le seuil.** On considère ici un réseau de contacts sociaux dans un lycée, enregistré par Salathé et collaborateurs [7]. En **A** on montre la fraction des nœuds infectés à l'équilibre (état endémique) en fonction de la transmissibilité, pour deux valeurs de probabilité de guérison (en rouge et bleu). Les flèches montrent la valeur théorique du seuil, calculée avec le propagateur d'infection. Les croix (MC) ont été obtenues de la solution numérique du processus de Markov (Equation 2 dans [68]), tandis que les points gris représentent le résultat des simulations microscopiques, dans lesquelles chaque nœud infectieux peut contaminer les nœuds sains avec lesquels il est en contact (ses voisins). En **B** on montre la valeur du seuil épidémique $\lambda_c(T)$ obtenu en considérant que le premiers T pas de temps. L'histogramme en gris est en échelle linéaire et montre le degré moyen associé au pas de temps T.

contact empiriques registrent seulement une partie limitée dans le temps des interactions, il est important d'évaluer l'impact d'une telle longueur temporelle sur l'estimation du seuil, et par conséquent notre capacité à évaluer la vulnérabilité du système . En [68] on explore cet aspect, en essayant d'identifier la période de prise de donnée minimale qui soit optimale pour caractériser le propagateur d'infection ; on en montre un exemple en Figure R2**B**.

En plus de la durée totale de la prise de données, le calcul du seuil permet d'évaluer l'impact d'autres caractéristiques des prises des données empiriques, y compris la résolution temporelle, que nous analysons dans [69].

La représentation multi-couches du réseau temporel, qui est le cœur de notre méthodologie de calcul du seuil, n'est possible que si on suppose que l'évolution du réseau se déroule en temps discret, ce qui représente une limite lorsque nous traitons des systèmes pour lesquels le temps est un paramètre continu. Pour résoudre ce problème, on montre que la limite continue du propagateur d'infection est obtenue en résolvant un système différentiel qui contient la matrice d'adjacence $A(t)$ du réseau temporel :

$$\frac{d}{dt}P(t) = P(t)\left[-m + lA(t)\right],\tag{3}$$

où l, m sont les taux de transmission et de guérison de la maladie. Une fois qu'on a trouvé le propagateur d'infection, le seuil peut être calculé avec son rayon spectral, comme dans le cas discret.

Les modèles épidémiques dont nous avons parlé jusqu'à maintenant ne prennent pas en compte une caractéristique commune à nombreuses maladies : la période de latence, c'est-à-dire le temps qui passe entre être infecté et devenir infectieux. Dans le cas des réseaux statiques la latence n'a pas d'impact sur le seuil épidémique [50], par contre cela n'est plus vrai pour les réseaux temporels, en raison de l'interaction entre les trois échelles de temps en jeu : évolution du réseau, durée d'infection, et période de latence. Le modèle SIS peut être modifié pour prendre en compte ce nouvel ingrédient, en devenant susceptible-exposé-infectieux-sensible (SEIS). La progression de la maladie est similaire au cas du SIS, sauf que les individus infectées ne sont pas immédiatement infectieux (E) ; ils deviennent infectieux avec une probabilité ϵ à chaque pas de temps.

Notre objectif est de calculer le propagateur d'infection du modèle SEIS sur réseaux temporels, et pour ce faire, nous utilisons à nouveau le formalisme des réseaux multi-couches. En particulier, on part du cas statique, et on démontre que le modèle SEIS sur un réseau A générique a le même seuil d'un modèle SIS sur un réseau à 2 couches. La structure spécifique du réseau 2-couches et les paramètres du modèle SIS permettant cette analogie sont décrits dans la Figure R3. Une fois démontré cette équivalence, la généralisation au cas temporel est obtenue en appliquant les résultats précédents [68,69]. On obtient ansi le propagateur d'infection $\hat{P}$ pour le modèle SEIS, en termes de matrices composées de quatre blocs $N \times N$:

$$\hat{P} = \prod_{t=1}^{T}\left(\begin{array}{c|c} 1-\mu & \epsilon \\ \hline \lambda A_t & 1-\epsilon \end{array}\right)\tag{4}$$
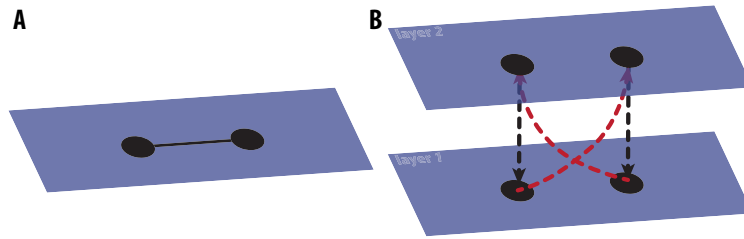
FIGURE R3: **Représentation de la structure à deux couches pour le calcul du seuil épidémique en incluant la latence. A** représente le réseau statique initial. **B** représente la projection multi-couches. Chaque nœud a une copie sur chaque couche, et la copie dans la couche 2 est connectée à la copie dans la couche 1 avec un lien direct (flèches noires pointillées). Chaque lien présent en **A** se traduit par deux liens directs de la couche 1 à la couche 2 (flèches rouges pointillées). À cette structure on couple un processus SIS, dans lequel les liens noirs ont une probabilité de transmission égale à ϵ, tandis que les rouges λ. De plus, la probabilité de guérison dans la couche 1 est μ, et dans la couche 2 est ϵ.

Réseaux de déplacements des bovins

Les maladies des animaux de ferme ont une grande pertinence dans le contexte de la santé publique, car elles menacent la santé et le bien-être animal, et ont un effet négatif sur l'économie, en termes de réduction de la productivité et de coûts de l'éradication [74, 75]. De plus, elles représentent un danger direct pour la santé humaine [76].

En raison de sa position géographique et son marché de bétail intégré, l'Europe est constamment menacée par divers agents pathogènes, y compris la fièvre aphteuse [74, 77, 78], la diarrhée virale bovine [79–81], la tuberculose bovine [82–84] et la brucellose [85–87] pour les bovins, la fièvre catarrhale du mouton pour tous les ruminants, la peste porcine classique [88] et africaine [89] pour les porcs. Bien que ces maladies ont des caractéristiques épidémiologiques très différentes les unes des autres, elles se propagent toutes, exclusivement ou en partie, grâce aux mouvements d'animaux entre les structures ; pour cette raison, l'étude des réseaux de mouvements est extrêmement importante. De nombreuses études ont déjà été publiées, sur des pays et espèces différents. Dans cette thèse, nous nous concentrons sur le marché des bovins, et développons une analyse comparative des mouvements dans trois pays européens : l'Italie, la Suède, la Hongrie, grâce à la coopération avec les institutions compétentes [1]. La Figure R4 montre la définition des réseaux, et leur représentation agrégée. Ces réseaux sont intrinsèquement temporels, parce que les mouvements sont enregistrés quotidiennement.

Nous avons observé que ces réseaux montrent une phénoménologie extrêmement riche et diversifiée en termes de propriétés globales, temporelles, et géographiques. En par-

---

1. Istituto Zooprofilattico Sperimentale "G. Caporale" (Italie), Linköping University (Suède), Hungarian National Food Chain Safety Office (Hongrie).
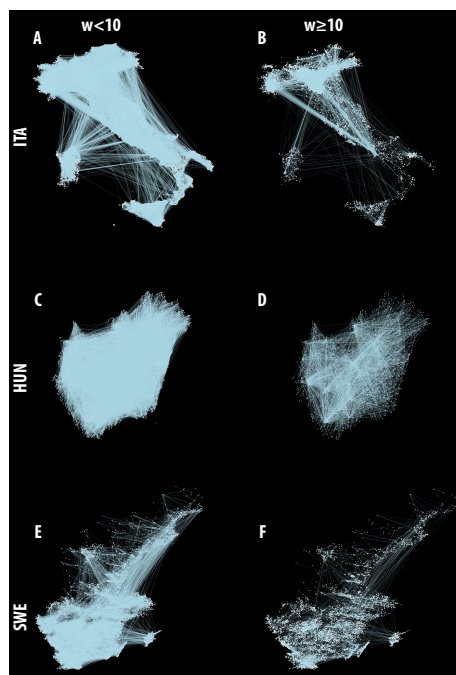
FIGURE R4: **Réseau bovin agrégé.** Représentation du réseau des mouvements des bovins, agrégé sur une période d'un an : 2008 pour l'Italie (**A**,**B**) et la Suède (**E**,**F**), 2010 pour l'Hongrie (**C**,**D**). Les nœuds (les structures qui participent au marché) sont placés en fonction de leur position géographique ; les lien représentent les déplacements de bovins d'un nœud à l'autre, composés de moins (**A**,**C**,**E**) ou plus de 10 animaux (**B**,**D**,**F**).

ticulier, les réseaux agrégés à échelles de temps différentes présentent des topologies hétérogènes dans les trois pays. En outre, l'échelle de temps qui domine l'évolution de chaque réseau est la semaine ; par contre les effets saisonniers tendent à être spécifiques à chaque pays. Les caractères spatiaux aussi ont tendance à varier d'un pays à l'autre, car ils répondent à la diversité climatique et géographique.

Analyse du risque

Le calcul du seuil épidémique peut être utilisé comme une estimation du risque, pour mesurer combien une population est vulnérable à l'introduction d'un agent pathogène spécifique. Pour ce faire, nous nous sommes concentrés sur deux cas : le réseau des mouvements de bovins italiens et le réseau de contacts sexuels humains (données en [34]). Pour les bovins, nous avons calculé le seuil des modèles SIS, SIR, SIRS, qui ensemble couvrent une bonne variété d'infections, et nous en avons étudié les variations dans le temps et l'espace. Nous avons observé une forte hétérogénéité spatiale, avec la partie nord de l'Italie extrêmement plus vulnérable que le Centre et le Sud, en raison des différences dans les types d'élevage, comme on avait déjà trouvé en étudiant ce réseau. De plus, on a trouvé une hétérogénéité temporelle probablement attribuable à des adaptations du réseau suite à des urgences sanitaires, ou à l'adoptions de nouveaux règlements.

Dans le contexte des contacts sexuels humains, nous avons étudié l'impact de la latence sur la vulnérabilité du réseau, en terme de différence du seuil épidémique entre le modèle sans latence (SIS), et avec latence (SEIS). Nous nous sommes concentrés sur la propagation de la syphilis à travers ce réseau, et nous avons observé comment, en explorant les paramètres épidémiques, la latence peut représenter soit un facteur de risque soit un facteur de protection.

Le calcul du risque avec le seuil épidémique a cependant une limitation due au fait que les données de contact ne sont guère disponibles en temps réel, et en plus dans de nombreux contextes – notamment pour les mouvements des bovins [38] – les mesures de centralité calculées sur les données passées ne sont pas représentatives des développements futurs. De plus, le seuil est une mesure de la vulnérabilité de l'ensemble de la population ; elle n'est capable de fournir des prévisions spécifiques sur la probabilité qu'a un certain nœud d'être infecté. Pour répondre à ces questions, nous avons développé dans [72] une méthodologie pour prédire le risque épidémique au niveau des nœuds, en n'utilisant que les données de contacts passées.

La quantité clé de cette méthode est ce que nous appelons *loyalty*, qui mesure la tendance de chaque nœud à rétablir des connexions toujours avec les mêmes nœuds, en opposition à la tendance à changer souvent de voisins. Si on considère les pas de temps t et $t+1$ d'un réseau temporel, et $\mathcal{V}_i^t$ l'ensemble des voisins du nœud i au temps t, la loyalty du nœud i entre les temps t et $t+1$ $(\theta_i^{t,t+1})$ est donc $\theta_i^{t,t+1} = \left|\mathcal{V}_i^t \cap \mathcal{V}_i^{t+1}\right| / \left|\mathcal{V}_i^t \cup \mathcal{V}_i^{t+1}\right|$. La loyalty prend valeurs dans $[0,1]$; $\theta = 0$ signifie qu'aucun voisin est maintenu, tandis que $\theta = 1$ indique que le nœud conserve exactement les mêmes voisins. Cette mesure trouve son inspiration dans le contexte des mouvements des bovins, comme une estimation de la tendance de l'éleveur à acheter du bétail chez les mêmes fournisseurs. En considérant

à la fois les mouvements bovins et les contacts sexuels humains, nous avons effectué des simulations épidémiques, en supposant un modèle susceptible-infecté, afin de modéliser les premiers stades d'une épidémie [72]. Nous sommes ensuite allés observer les différences dans la façon de propagation de l'épidémie, causée par les changements temporels dans le réseau. Nous avons constaté qu'un nœud qui est infecté lors d'une flambée épidémique au temps t, sera infecté par la même épidémie simulée au temps $t+1$ avec une probabilité corrélée positivement avec sa loyalty. Ce résultat, observé numériquement dans les deux réseaux, est à la base d'une méthode permettant de prédire le risque que qu'un nœud a d'être infecté dans le futur par une épidémie spécifique.

Enfin, pour essayer de mieux comprendre les mécanismes responsables de la performance de cette méthode, nous avons développé un modèle synthétique de réseau qui garde uniquement les propriétés topologiques de base communes aux deux réseaux réels : hétérogénéité de dégré à tout niveau d'agrégation temporelle, et un niveau de mémoire dans l'évolution, en termes de distribution de loyalty, réglable par un paramètre. Nous avons observé que ces ingrédients sont suffisants pour assurer la précision de nos prévisions, et que la mémoire est ce qui règle la puissance prédictive, c'est-à-dire la fraction de nœuds infectés par l'épidémie pour lesquels on a pu produire des prévisions de risque.
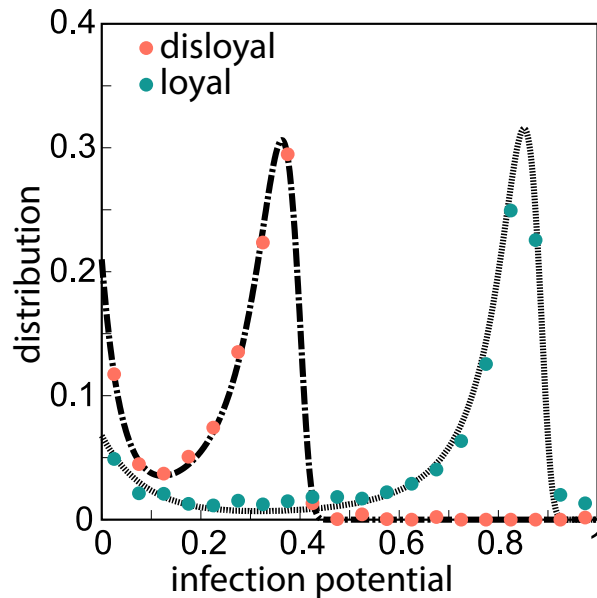

FIGURE R5: **potentiel de propagation pour le réseau bovin italien.** Les nœuds sont classifiés comme *disloyal* ($\theta \leq 0.1$, orange) ou *loyal* ($\theta > 0.1$, vert). On montre les distributions de la probabilité que un nœud *disloyal* (*loyal*), infecté pendant la simulation pour l'année t, soit infecté par la même épidémie pendant l'année $t+1$ (potentiel de propagation). On observe que le potentiel de propagation d'un nœud *loyal* est en moyenne 2.5 plus grand que celui d'un nœud *disloyal*.

Conclusions

Dans cette thèse, on a développé de nouvelles méthodes pour estimer et prédire le risque associé à l'émergence d'un nouveau pathogène, à la fois à l'échelle de la population et à celle de l'hôte, et en utilisant des outils à la fois analytiques et numériques. Nous avons testé ces méthodes dans des milieux différents, tous très pertinents dans le contexte de la santé publique, avec un accent sur les réseaux des mouvements de bovins dans de différents pays européens, pour lesquels nous avons proposé une analyse des caractéristiques topologiques, spatiales et temporelles.

Références

[1] Norman Bailey. *The mathematical theory of infectious diseases and its applications.* Griffin, London, 1975.

[2] R M Anderson and Robert M May. *Infectious Diseases of Humans : Dynamics and Control.* Oxford University Press, 1992.

[3] Daryl Daley. *Epidemic modelling : an introduction.* Cambridge University Press, Cambridge New York, 1999.

[4] Matt J. Keeling and P. Rohani. Modeling Infectious Diseases in Humans and Animals. *Princeton University Press*, 2007.

[5] Emilia Vynnycky. *An introduction to infectious disease modelling.* Oxford University Press, Oxford, 2010.

[6] Ciro Cattuto, Wouter den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE*, 5(7) :e11596, 2010.

[7] Marcel Salathé and James H Jones. Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Comput Biol*, 6(4) :e1000736, 2010.

[8] Mohammad S Hashemian, Kevin G Stanley, and Nathaniel Osgood. FLUNET : Automated Tracking of Contacts During Flu Season. In *WiOpt'10 : Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, pages 557–562, Avignon, France, May 2010.

[9] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W Feldman, and James H Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51) :22020–22025, 2010.

[10] Lorenzo Isella, Mariateresa Romano, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Wouter Van den Broeck, Francesco Gesualdo, Elisabetta Pandolfi, Lucilla Ravà, Caterina Rizzo, and Alberto Eugenio Tozzi. Close Encounters in a Pediatric Ward : Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors. *PLOS ONE*, 6(2) :e17144, 2011.

[11] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd ? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1) :166–180, 2011.

[12] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems. High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. *PLOS ONE*, 6(8) :e23176, 2011.

[13] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Regis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, and Philippe Vanhems. Simulation of an SEIR Infectious Disease Model on the Dynamic Contact Network of Conference Attendees. *BMC Medicine*, 9(87), July 2011.

[14] Thomas Hornbeck, David Naylor, Alberto M Segre, Geb Thomas, Ted Herman, and Philip M Polgreen. Using Sensor Networks to Study the Effect of Peripatetic Healthcare Workers on the Spread of Hospital-Associated Infections. *The Journal of Infectious Diseases*, 206(10) :1549–1557, November 2012.

[15] Wouter Van den Broeck, Marco Quaggiotto, Lorenzo Isella, Alain Barrat, and Ciro Cattuto. The making of Sixty-Nine Days Of Close Encounters At The Science Gallery. *Leonardo*, 45(3) :285, May 2012.

[16] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, and Nicolas Voirin. Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. *PLoS ONE*, 8(9) :e73970, 2013.

[17] Julie Fournet and Alain Barrat. Contact Patterns among High School Students. *PLoS ONE*, 9(9) :e107878, 2014.

[18] Nicolas Voirin, Cécile Payet, Alain Barrat, Ciro Cattuto, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, Jean-Sébastien Casalegno, Bruno Lina, and Philippe Vanhems. Combining High-Resolution Contact Data with Virological Data to Investigate Influenza Transmission in a Tertiary Care Hospital. *Infection Control & Hospital Epidemiology*, 36(03) :254–260, 2015.

[19] Thomas Obadia, Romain Silhol, Lulla Opatowski, Laura Temime, Judith Legrand, Anne C M Thiébaut, Jean-Louis Herrmann, Éric Fleury, Didier Guillemot, and Pierre-Yves Boëlle. Detailed Contact Data and the Dissemination of Staphylococcus aureus in Hospitals. *PLoS Comput Biol*, 11(3) :e1004170, 2015.

[20] Marta C González, Cesar A Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196) :779–782, June 2008.

[21] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of Predictability in Human Mobility. *Science*, 327(5968) :1018–1021, 2010.

[22] Gautier Krings, Márton Karsai, Sebastian Bernhardsson, Vincent D Blondel, and Jari Saramäki. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1(1) :1–16, 2012.

[23] Mikko Kivelä, Raj Kumar Pan, Kimmo Kaski, János Kertész, Jari Saramäki, and Márton Karsai. Multiscale analysis of spreading in a large communication network. *Journal of Statistical Mechanics : Theory and Experiment*, 2012(03) :P03005, 2012.

[24] Lauri Kovanen, Kimmo Kaski, János Kertész, and Jari Saramäki. Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proceedings of the National Academy of Sciences*, 110(45) :18070–18075, 2013.

[25] Zhi-Qiang Jiang, Wen-Jie Xie, Ming-Xia Li, Boris Podobnik, Wei-Xing Zhou, and H Eugene Stanley. Calling patterns in human communication dynamics. *Proceedings of the National Academy of Sciences*, 110(5) :1600–1605, 2013.

[26] Ming-Xia Li, Vasyl Palchykov, Zhi-Qiang Jiang, Kimmo Kaski, János Kertész, Salvatore Miccichè, Michele Tumminello, Wei-Xing Zhou, and Rosario N Mantegna. Statistically validated mobile communication networks : the evolution of motifs in European and Chinese data. *New Journal of Physics*, 16(8) :83038, 2014.

[27] Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent D Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza. On the Use of Human Mobility Proxies for Modeling Epidemics. *PLoS Comput Biol*, 10(7) :e1003716, 2014.

[28] Pablo Kaluza, Andrea Kölzsch, Michael T Gastner, and Bernd Blasius. The complex network of global cargo ship movements. *Journal of The Royal Society Interface*, January 2010.

[29] Lijun Sun, Kay W Axhausen, Der-Horng Lee, and Xianfeng Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34) :13774–13779, 2013.

[30] Pierre Borgnat, Céline Robardet, Patrice Abry, Patrick Flandrin, Jean-Baptiste Rouquier, and Nicolas Tremblay. A Dynamical Network View of Lyon's V\'elo'v Shared Bicycle System. In Animesh Mukherjee, Monojit Choudhury, Fernando Peruani, Niloy Ganguly, and Bivas Mitra, editors, *Dynamics On and Of Complex Networks, Volume 2 SE - 13*, Modeling and Simulation in Science, Engineering and Technology, pages 267–284. Springer New York, 2013.

[31] Martin Rosvall, Alcides V Esquivel, Andrea Lancichinetti, Jevin D West, and Renaud Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nat Commun*, 5, August 2014.

[32] Ingo Scholtes, Nicolas Wider, René Pfitzner, Antonios Garas, Claudio J Tessone, and Frank Schweitzer. Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nat Commun*, 5, September 2014.

[33] Ken T D Eames and Matt J. Keeling. Monogamous networks and the spread of sexually transmitted diseases. *Mathematical biosciences*, 189(2) :115–30, June 2004.

[34] Luis E C Rocha, Fredrik Liljeros, and Petter Holme. Information dynamics shape the sexual networks of Internet-mediated prostitution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13) :5706–5711, 2010.

[35] R R Kao, Leon Danon, D M Green, and Istvan Z Kiss. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proceedings of the Royal Society of London B : Biological Sciences*, 273(1597) :1999–2007, August 2006.

[36] Matthew C. Vernon and Matt J. Keeling. Representing the UK's cattle herd as static and dynamic networks. *Proceedings of the Royal Society of London B : Biological Sciences*, 276(1656) :469–476, February 2009.

[37] Tom Lindström, Scott A Sisson, Susanna Stenberg Lewerin, and Uno Wennergren. Estimating animal movement contacts between holdings of different production types. *Preventive veterinary medicine*, 95(1-2) :23–31, June 2010.

[38] Paolo Bajardi, Alain Barrat, Fabrizio Natale, Lara Savini, and Vittoria Colizza. Dynamical Patterns of Cattle Trade Movements. *PLoS ONE*, 6(5) :e19869, 2011.

[39] Mario Konschake, Hartmut H K Lentz, Franz J Conraths, Philipp Hövel, and Thomas Selhorst. On the Robustness of In- and Out-Components in a Temporal Network. *PLoS ONE*, 8(2) :e55223, 2013.

[40] Bhagat Lal Dutta, Pauline Ezanno, and Elisabeta Vergu. Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. *Preventive veterinary medicine*, 117(1) :79–94, November 2014.

[41] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[42] S N Dorogovtsev, A V Goltsev, and J F F Mendes. Critical phenomena in complex networks. *Rev. Mod. Phys.*, 80(4) :1275–1335, October 2008.

[43] Carter T Butts. Revisiting the Foundations of Network Analysis. *Science*, 325(5939) :414–416, 2009.

[44] Alessandro Vespignani. Predicting the Behavior of Techno-Social Systems. *Science*, 325(5939) :425–428, 2009.

[45] M E J Newman. *Networks : an introduction*. Oxford University Press, Oxford New York, 2010.

[46] Chapter 12 - An Overview of Social Networks and Economic Applications. In *Handbook of Social Economics*, volume 1, pages 511–585. 2011.

[47] Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nat Phys*, 8(1) :32–39, January 2012.

[48] Albert-László Barabási. *Network science*. Cambdridge University Press, Cambridge UK, 2015.

[49] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14) :3200–3203, 2001.

[50] M E J Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66(1) :16128, 2002.

[51] Uwe Naumann. *Combinatorial scientific computing*. CRC Press, Boca Raton, 2012.

[52] Yang Wang, D Chakrabarti, Chenxi Wang, and C Faloutsos. Epidemic spreading in real networks : an eigenvalue viewpoint. In *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, pages 25–34, 2003.

[53] Sergio Gómez, Alexandre Arenas, J Borge-Holthoefer, Sandro Meloni, and Yamir Moreno. Discrete time Markov chain approach to contact-based disease spreading in complex networks. *Europhysics Letters*, 89(3) :38009, 2010.

[54] Alexei Vazquez, Balázs Rácz, András Lukács, and Albert-László Barabási. Impact of Non-Poissonian Activity Patterns on Spreading Processes. *Phys. Rev. Lett.*, 98(15) :158702, April 2007.

[55] José Luis Iribarren and Esteban Moro. Impact of Human Activity Patterns on the Dynamics of Information Diffusion. *Phys. Rev. Lett.*, 103(3) :38702, July 2009.

[56] Márton Karsai, Mikko Kivelä, R K Pan, K Kaski, J Kertész, Albert-László Barabási, and Jari Saramäki. Small but slow world : How network topology and burstiness slow down spreading. *Phys. Rev. E*, 83(2) :25102, February 2011.

[57] Giovanna Miritello, Esteban Moro, and Rubén Lara. Dynamical strength of social ties in information spreading. *Phys. Rev. E*, 83(4) :45102, April 2011.

[58] Luis E C Rocha and Vincent D Blondel. Bursts of Vertex Activation and Epidemics in Evolving Networks. *PLoS Comput Biol*, 9(3) :e1002974, 2013.

[59] Luca Ferreri, Paolo Bajardi, Mario Giacobini, Silvia Perazzo, and Ezio Venturino. Interplay of network dynamics and heterogeneity of ties on spreading dynamics. *Phys. Rev. E*, 90(1) :12812, July 2014.

[60] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3) :97–125, October 2012.

[61] Petter Holme and Jari Saramäki, editors. *Temporal networks*. Springer, Berlin New York, 2013.

[62] Petter Holme. Modern temporal network theory : A colloquium. *arXiv*, 2015.

[63] Thilo Gross, Carlos J Dommar D'Lima, and Bernd Blasius. Epidemic Dynamics on an Adaptive Network. *Phys. Rev. Lett.*, 96(20) :208701, May 2006.

[64] Erik Volz and Lauren Ancel Meyers. Epidemic thresholds in dynamic contact networks. *Journal of The Royal Society Interface*, 6(32) :233–241, March 2009.

[65] Michael Taylor, Timothy J Taylor, and Istvan Z Kiss. Epidemic threshold and control in a dynamic network. *Phys. Rev. E*, 85(1) :16103, January 2012.

[66] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Sci. Rep.*, 2, June 2012.

[67] Suyu Liu, Nicola Perra, Márton Karsai, and Alessandro Vespignani. Controlling Contagion Processes in Activity Driven Networks. *Phys. Rev. Lett.*, 112(11) :118702, March 2014.

[68] Eugenio Valdano, Luca Ferreri, Chiara Poletto, and Vittoria Colizza. Analytical Computation of the Epidemic Threshold on Temporal Networks. *Phys. Rev. X*, 5(2) :21005, April 2015.

[69] Eugenio Valdano, Chiara Poletto, and Vittoria Colizza. Infection propagator approach to compute epidemic thresholds on temporal networks : impact of immunity and of limited temporal resolution. *arXiv*, 2015.

[70] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A Porter, Sergio Gómez, and Alexandre Arenas. Mathematical Formulation of Multilayer Networks. *Phys. Rev. X*, 3(4) :41022, December 2013.

[71] Mikko Kivelä, Alexandre Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of Complex Networks*, 2(3) :203–271, September 2014.

[72] Eugenio Valdano, Chiara Poletto, Armando Giovannini, Diana Palma, Lara Savini, and Vittoria Colizza. Predicting Epidemic Risk from Past Temporal Contact Data. *PLoS Comput Biol*, 11(3) :e1004152, 2015.

[73] Piet Van Mieghem. The N-intertwined SIS Epidemic Network Model. *Computing*, 93(2-4) :147–169, 2011.

[74] Iain Anderson. Foot and Mouth Disease 2001 : Lessons to be Learned Inquiry Report. *The Stationery Office*, (July), 2002.

[75] T. E. Carpenter, J. M. O'Brien, A. D. Hagerman, and B. A. McCarl. Epidemic and Economic Impacts of Delayed Detection of Foot-And-Mouth Disease : A Case Study of a Simulated Outbreak in California. *Journal of Veterinary Diagnostic Investigation*, 23(1) :26–33, January 2011.

[76] L H Taylor, S M Latham, and Mark E J Woolhouse. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B : Biological Sciences*, 356(1411) :983–989, July 2001.

[77] Matt J. Keeling, Mark E J Woolhouse, Darren J Shaw, Louise Matthews, Margo Chase-Topping, Dan T Haydon, Stephen J Cornell, Jens Kappey, John Wilesmith, and Bryan T Grenfell. Dynamics of the 2001 UK Foot and Mouth Epidemic : Stochastic Dispersal in a Heterogeneous Landscape. *Science*, 294(5543) :813–817, 2001.

[78] Matt J. Keeling. Models of foot-and-mouth disease. *Proceedings of the Royal Society of London B : Biological Sciences*, 272(1569) :1195–1202, June 2005.

[79] A-F Viet, C Fourichon, and H Seegers. Review and critical discussion of assumptions and modelling options to study the spread of the bovine viral diarrhoea virus (BVDV) within a cattle herd. *Epidemiology and infection*, 135(5) :706–721, 2007.

[80] Aurélie Courcoul and Pauline Ezanno. Modelling the spread of Bovine Viral Diarrhoea Virus (BVDV) in a managed metapopulation of cattle herds. *Veterinary Microbiology*, 142(1-2) :119–128, 2010.

[81] Mark Tinsley, Fraser I Lewis, and Franz Brülisauer. Network modeling of BVD transmission. *Veterinary research*, 43(1) :11, January 2012.

[82] L.M. O'Reilly and C.J. Daborn. The epidemiology of Mycobacterium bovis infections in animals and man : A review. *Tubercle and Lung Disease*, 76 :1–46, August 1995.

[83] Robin A. Skuce, Adrian R. Allen, and Stanley W. J. McDowell. Herd-level risk factors for bovine tuberculosis : A literature review. *Veterinary Medicine International*, 2012 :621210, 2012.

[84] Ellen Brooks-Pollock, Gareth O. Roberts, and Matt J. Keeling. A dynamic model of bovine tuberculosis spread and control in Great Britain. *Nature*, 511(7508) :228–231, July 2014.

[85] V Taleski, L Zerva, T Kantardjiev, Z Cvetnic, M Erski-Biljic, B Nikolovski, J Bosnjakovski, V Katalinic-Jankovic, A Panteliadou, S Stojkoski, and T Kirandziski. An overview of the epidemiology and epizootology of brucellosis in selected countries of Central and Southeast Europe. *Veterinary Microbiology*, 90(1-4) :147–155, December 2002.

[86] T England, L Kelly, R D Jones, A MacMillan, and M Wooldridge. A simulation model of brucellosis spread in British cattle under several testing regimes. *Preventive veterinary medicine*, 63(1-2) :63–73, April 2004.

[87] E Díaz Aparicio. Epidemiology of brucellosis in domestic animals caused by Brucella melitensis, Brucella suis and Brucella abortus. *Revue scientifique et technique (International Office of Epizootics)*, 32(1) :43–51, 53–60, 2013.

[88] V Moennig, G Floegel-Niesmann, and I Greiser-Wilke. Clinical Signs and Epidemiology of Classical Swine Fever : A Review of New Knowledge. *The Veterinary Journal*, 165(1) :11–20, January 2003.

[89] S Costard, L Mur, J Lubroth, J M Sanchez-Vizcaino, and D U Pfeiffer. Epidemiology of African swine fever virus. *Virus research*, 173(1) :191–7, April 2013.

PHD THESIS

Computing the vulnerability of time-evolving networks to infections

Eugenio Valdano

Sorbonne Universités, UPMC Univ Paris 06, INSERM, Institut Pierre Louis
d'Épidémiologie et de Santé Publique (IPLESP UMRS 1136), F75012, Paris, France.

Research articles published as first author contained in this thesis

- Eugenio Valdano, Luca Ferreri, Chiara Poletto, and Vittoria Colizza.
  Analytical Computation of the Epidemic Threshold on Temporal Networks.
  Physical Review X, 5(2):21005, 2015.
  Cited in this thesis as [1].

- Eugenio Valdano, Chiara Poletto, Armando Giovannini, Diana Palma, Lara Savini, and Vittoria Colizza.
  Predicting Epidemic Risk from Past Temporal Contact Data.
  PLoS Computational Biology, 11(3):e1004152, 2015.
  Cited in this thesis as [2].

Research articles submitted as first author contained in this thesis

- Eugenio Valdano, Chiara Poletto, and Vittoria Colizza.
  Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution.
  2015. - *under review*.
  Cited in this thesis as [3].

Quare habe tibi quicquid hoc libelli
qualecumque; quod, patrona virgo,
plus uno maneat perenne saeclo.

Catullus, Carmen I, 8-10

Abstract

Infectious disease modeling represents a powerful tool for assessing the vulnerability of a population to the introduction of a new infectious pathogen. The increased availability of highly resolved data tracking host interactions is making epidemic models potentially increasingly accurate. Integrating into them all the features emerging from these data, however, still represents a challenge. In particular, the interaction between disease dynamics and the time evolution of contact structures has been shown to impact the way pathogens spread. Specifically, it changes the conditions that lead to the wide-spreading regime, as encoded in epidemic threshold, which is the critical transmissibility value above which the epidemic breaks out. Up to now, through the formalism of temporal networks, researchers have characterized the epidemic threshold on time evolving contact structures only through numerical approaches or in specific settings. Using a multilayer formalism, we analytically compute the epidemic threshold on a generic temporal network, accounting for several different disease features. We use this methodology to assess the impact of time resolution and network duration on the estimation of the threshold, given the importance that these features have in setting up an optimal data collection strategy. We introduce several epidemiologically relevant datasets, and in particular we analyze the networks of cattle trade movements, highlighting the features that can influence disease spreading. Then we use the analytical tools we have developed to assess the global vulnerability of different systems to pathogen introduction, focusing on Italian cattle trade movements, and on a network of human sexual contacts. Data collection strategies often inform us only about past network configurations, and that limits our prediction capabilities. We face this by developing a data-driven methodology for predicting targeted epidemic that relies only past contact data, and apply to real and synthetic networks. Our work provides new methodologies for assessing and predicting the risk associated to an emerging pathogen, both at the population scale and targeting specific hosts. We develop and test them in a wide variety of contexts, making our results widely applicable.

Contents

Introduction

Infectious diseases represent a major burden on welfare and society. They directly threaten human health, and impact economy and development. In developing countries, they represent the top cause of death [4], and pose a particularly heavy burden on child health, with pneumonia, diarrhea and malaria together accounting for more than 60% percent of child deaths [5]. Even in the developed World, where most deaths are due to non-infectious diseases, the situation may worsen in the near future. Bacteria strains are developing antibiotic resistance at a quicker pace than we can come up with new drugs [6], and vector-borne diseases (for instance, Dengue Fever [7]) are now reaching areas in which they were previously absent, as climate change impacts vector ecology. Finally, the ever more globalized World we live in is prone to breakouts of pathogens with pandemic potential, like SARS (2003) [8–10], H1N1 flu (2009) [11–15], or, more recently MERS CoV [16–19] and Ebola [20–23]. In the fight against infectious diseases, mathematical models have become an important tool, as they provide tools to reduce the number of infections, optimally allocate limited resources, react promptly to emergencies, and implement targeted containment strategies. Ever since their introduction in 1927 by Kermack & McHendrick [24–26], these models are based on the assumption that disease transmission and progression can be translated into a relative simple set of mechanistic rules that can be adapted to pathogens with very diverse pathophysiology, clinical symptoms, and causative agents (bacteria, viruses, etc.). The epidemic is then an emerging collective behavior of the "microscopic" interactions among hosts [27–31]. This framework has allowed researchers to apply and develop several tools and techniques borrowed from mathematics and statistical physics. Up until two decades ago, however, the use of mathematical models in public health was limited by the lack of data concerning human interactions. Simplified and coarse-grained assumptions limited model applicability to real scenarios. The picture has dramatically changed in the last years, with the outbreak of *data science*. The development of both new hardware and software technologies has made it possible to track contacts and transports relevant for

the spread of diseases in an extremely accurate and almost unsupervised way. We now have detailed records of how people interact in different settings, and at different scales, from face-to-face proximity encounters [32–45], to mobility patterns [46–58], to sexual contacts [59, 60]. Detailed data do not concern only human activities: we can now keep track of livestock displacements between farms, which are spreading routes for many diseases threatening animal health, economy, and ultimately human health [61–66].

This "data deluge"[1] has radically transformed infectious disease modeling, proving to be both a huge resource and a great challenge. High resolution data have made it possible to model entire populations down to single individuals [67–73], providing tailored real-time predictions of epidemic outbreaks [15, 74, 75]. These simulations schemes are effective and perform well in specific settings, but cannot provide a general understanding of the unfolding of epidemic processes. New theoretical tools have been developed in order to deal with the complexity of interaction structures, treating diseases as dynamic processes on networks [76–83]. Networks have indeed become a common and successful tool to model populations in terms of nodes (hosts) and links (interactions among hosts) of a graph [84–86]. This has made it possible to use all the tools already developed in statistical physics, graph theory, and dynamical systems, to the context of infectious diseases.

An important feature emerging from this huge amount of data is that contacts are not fixed in time during the spread of the disease, but evolve in time. The complex interplay between network evolution and disease diffusion has been proven to impact the outcome of the epidemic process, in several different contexts [39, 87–92]. This is due to time correlations between contacts, that determine the shape and amount of routes along which the disease can spread, and make traditional risk proxies, developed for static contact patterns, insufficient to characterize this new phenomenology [64, 93].

As a result, a methodology that aims at assessing the threat a specific pathogen poses to a population must account for the temporal evolution of contacts within that population. This calls for new theoretical tools to handle the interaction between the dynamics of network evolution, and the spread of the pathogen. Temporal networks represent an effective framework to model time-evolving contact patterns [94–96], but their interplay with spreading dynamics has so far been investigated only though numerical approaches, or in controlled settings [92, 97–101]. In the first part of our work we focus on assessing

---

[1]The data deluge, *The Economist*, 25th February 2010.

the vulnerability of a system to a disease, in term of its epidemic threshold, i.e., the critical value of disease transmissibility above which the epidemic breaks out. We introduce a new way of representing the unfolding of a disease on a time-evolving contact structure in terms of a multilayer network [102, 103], and use it to analytically compute the epidemic threshold on a generic temporal network. Our computation can account for various disease features, like immunity, waning of immunity and latency period. We then focus on particular datasets that carry a notable epidemiological relevance. Specifically, we provide a comparative analysis of the network of cattle trade movements in three European countries, highlighting the structural features that can impact the spread of an epidemic. In the last part of the thesis we develop a risk assessment analysis on these specific datasets, focusing on particular classes of diseases. Thanks to our theoretical findings, we can use the threshold as a measure of global network risk. Then, we focus on a more targeted, host-centered, risk analysis. Given that, for many systems, contacts are not reported in real time, the datasets that are commonly available inform us on the structure of contacts in the past, and that limits our prediction capabilities. We thus develop a data-driven methodology that, in case of a future epidemic outbreak of a certain pathogen, is able to provide the risk a particular host will be hit, by relying only to past contact data. We apply this methodology to different real scenarios, and develop a generative network model to test which structural features of contacts determine the performance of our prediction.

The goal of this thesis is thus to present new tools to assess and predict the risk associated to the introduction of a pathogen onto a population, accounting for the dynamic evolution of the pattern of contacts among hosts. We do that through new analytical tools that are able to handle the interplay between network and disease dynamics, and through data-driven analyses and models. The variety of datasets and disease models we consider, albeit being a simplification of the actual progression of real pathogen, makes our findings applicable to various contexts and scenarios.

In Chapter 1 we introduce the basics of epidemic modeling and networks. We also define the epidemic threshold, and show how to compute it in different settings. In Chapter 2 we define temporal networks, their epidemiological relevance, and their representations. We also introduce some datasets that will be used in the following chapters. We then describe some generative models and reference models for temporal networks. In Chapter 3 we develop our methodology for computing the epidemic threshold on temporal networks. We extend it to generic networks, and different disease models.

We also assess the impact of temporal aggregation of contact data, and the length of the period of data collection in empirical settings. This chapter includes two research articles: *Analytical computation of the epidemic threshold on temporal networks*, published on *Physical Review X* [1], and *Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution*, currently under review [3]. In Chapter 4 we provide a comparative analysis of cattle displacement networks in three European countries. In Chapter 5 we compute epidemic risk in real settings, cattle trade movements (Chapter 4) and sexual contact network (Chapter 2 and [60]). In this chapter we develop and apply our framework for computing node epidemic risk from past contact data, contained in our article *Predicting Epidemic Risk from Past Temporal Contact Data*, published in *PLoS Computational Biology* [2].

1. Epidemics on static networks

Infectious diseases vary widely in their pathophysiology, clinical symptoms and etiology, resulting in diverse progression and transmission patterns. Viral diseases, like influenza, measles, chicken pox, usually confer permanent immunity after recovery, while bacterial diseases (tuberculosis, syphilis) allow multiple re-infections of the same host. Many diseases are transmitted by direct contact between hosts, while others require vectors, such as malaria or bluetongue. Some, like cholera, require the ingestion of contaminated water and food. Such diversity calls for modeling approaches that are general and versatile enough to be adapted to each specific ailment, and still be a realistic description of its epidemiological features [104]. This is commonly achieved through *compartmental models*, which will be the first argument of this chapter. We will then explain how to account for the contact structure of the host population, and how compartmental models are adapted to complex population structures. Finally, we will investigate what are the conditions that, following a pathogen introduction in a susceptible population, lead to an epidemic outbreak. We will formalize this concept in terms of *epidemic threshold*, and show how to compute it for several contact structures.

1.1. Compartmental models

In principle, the mathematical description of an infectious disease can explore different scales, from describing pathogen population dynamics within a host, to ecological models that predict incidence in a population. We are interested in modeling the spread of a disease on a host population, so we consider the health status of a host as the elementary building block of our modeling approach. In particular, we assume that the health status can be described as a discrete set of states (compartments). The two compartments that characterize every epidemic model are *susceptible* (S) to being infected, and *infectious* (I), i.e., infected and contagious. Another compartment commonly used is *recovered* (R), i.e., immune to the disease. Hosts in the population are divided into compartments, ac-

cording to their health status [28, 105], and disease progression is modeled in terms of a set of interaction and transition rules among compartments. The simplest conceivable model is the susceptible-infected (SI). In the SI model, infectious hosts infect their susceptible neighbors with probability λ, and no recovery is possible. Despite its extreme simplicity, the SI model can be used, for instance, to approximate the initial stage of an outbreak (Section 5.3 and [2]). Other widely used models are the susceptible-infectious-susceptible (SIS) and susceptible-infectious-recovered (SIR) models. In both models, infectious hosts infect their susceptible neighbors with probability λ. In addition, they recover with probability μ, going either back to the susceptible state (SIS), or to the recovered state (SIR). In the SIR model recovered hosts are immune to re-infection. Probabilities are usually defined in discrete time; when time is continuous we have rates instead of probabilities, i.e., probabilities per infinitesimal time interval (see also Section 3.3 and [3]). The success of these two models is due to two aspects. Firstly their simplicity and their small number of free parameters make them analytically treatable, and their results provide a general understanding of infection dynamics. Secondly, they have proven adequate to model a wide range of diseases under simplifying assumptions. SIS model applies quite well to bacterial infections (see, for instance, hospital acquired MRSA [45], or STD's [91]), while SIR model works well with viral diseases, like flu, chicken pox [28], cattle Foot-and-Mouth [106]. More elaborate compartmental models account of more complex infection dynamics. Among them, we describe two that we will use in this thesis. The first is the SIRS model. Its progression is analogous to SIR model, but recovered individuals lose their immunity with probability ω, going back to susceptible state. SIR model can indeed be seen as a particular case of SIRS, when $\omega = 0$. The last model we use is the SEIS model. Many ailments are characterized by a latency period, which is a time lag between being infected, and becoming infectious. We can take this feature into account by adding a new compartment to the model, E (Exposed). A susceptible agent is infected by an infectious neighbor, and becomes exposed. It then turns infectious with a probability ϵ. From then, the progression is the same as the SIS model. Figure 1.1 schematically summarizes all the models we use in this thesis. For a detailed description of compartmental models we refer to traditional textbooks [27–31].
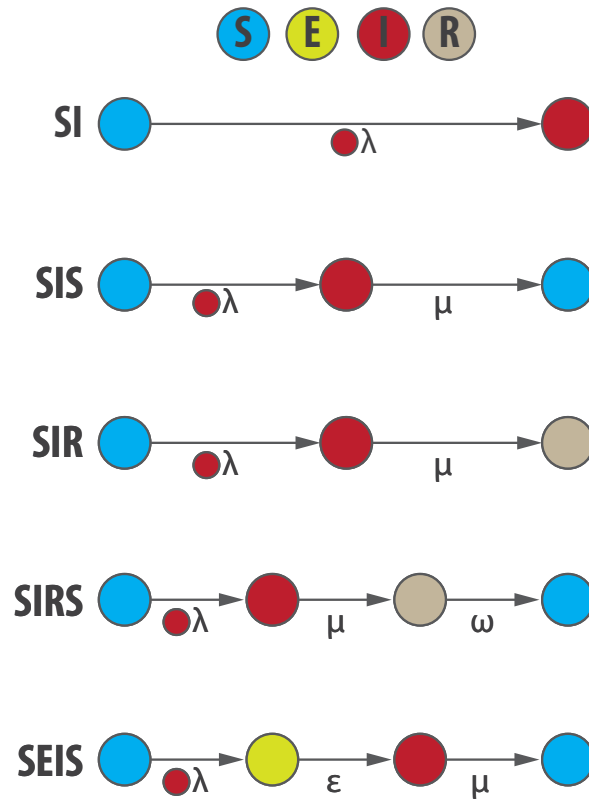
Figure 1.1.: **Compartmental models.** The first row describes the four compartments used. Susceptible is in blue, infectious in red, recovered in gray, and exposed in yellow. Dynamics of SI, SIS, SIR, SIRS and SEIS is then described. Arrows showing transitions from one compartment to the other are pictured with the respective probability. The red (infectious) smaller point close to λ indicates that infection requires interaction with an infectious agent. All other processes are spontaneous one-body transitions.

1.2. Homogeneous mixing

In general, the choice of compartmental model, and the values of the parameters, are informed by the epidemiology of the specific disease under study. The goal of infectious disease modeling is to uncover and characterize the macroscopic behavior emerging from applying such model to a population interacting in a certain way. In order to do that, it is necessary to have information about the structure of contacts along which the disease spreads.

The simplest framework is *homogeneous mixing*, which assumes no specific contact structure. Every host in the population has the same probability of being in contact with any other host. Despite its simplicity, homogeneous mixing still represents a successful approximation in many contexts, especially within patches of meta-population models [15, 107–111].

The traditional way to couple compartmental models to homogeneous mixing is represented by deterministic differential equations, interpreting pathogen spread as a reaction process. These equations compute the number of hosts in a compartment as a function of time. Let $S(t), I(t), R(t)$ be the number of susceptible (infectious, recovered) individuals at time t. We call N the population size, and k the number of contacts every host establishes. The equations of the SIS model are

$$\begin{cases} \dot{S} = -\frac{\lambda k}{N} SI + \mu I; \\ \dot{I} = \frac{\lambda k}{N} SI - \mu I. \end{cases} \tag{1.1}$$

For the SIR model are instead

$$\begin{cases} \dot{S} = -\frac{\lambda k}{N} SI; \\ \dot{I} = \frac{\lambda k}{N} SI - \mu I; \\ \dot{R} = \mu I. \end{cases} \tag{1.2}$$

Last line in both Equation 1.1 and Equation 1.2 is redundant if we assume a fixed population ($\dot{N} \equiv 0$).

Deterministic differential equations, however, have two important limitations. They treat the number of infected agents as a continuous variable, and they do not account for stochastic effects. In other words, they do not consider hosts as individual entities, which have a certain probability of transmitting to a contact, or change infection status. As a result, differential equations produce accurate predictions only in large populations (in

the limit $N \to \infty$). Real populations, however, often cannot be treated as infinite, and stochastic effects kick in. They not only quantitatively change the predictions, but have conceptual implications in defining what we mean by the persistence of an epidemic. One way to account for stochastic effects in the framework of homogeneous mixing is through *branching processes* [112–114]. In a SIS branching process, any infectious host in a given generation produces a random number of infectious hosts, sampled from a given distribution, which make up the next generation. In practice we seed a certain number of infectious hosts (first generation), they then will produce a certain number of infectious (second generation), and so on.

1.3. Networks

Epidemiologists have been reconstructing who had been in contact with whom during a particular time frame through surveys and questionnaires, asking, for instance, to list list all people you had met that particular day. This strategy has been used for different types of contacts, like face-to-face proximity [115–120], sexual encounters [121], and needle share among injection drug users [122–124]. These data allowed researchers to uncover features of human interactions, which prompted the need of going beyond the homogeneous mixing assumption. Researchers found out that these these complex contact structures could be represented as *networks* [76, 80, 82, 83, 125, 126].

A network is a representation of an interacting population, in terms of a mathematical entity, the *graph* [84–86, 127, 128]. A graph is composed of a set of nodes (vertices), and links (edges) that connect pairs of nodes. In our context, nodes represent the hosts of our population, and links represent the interactions relevant to the spread of the disease. The first elementary concept that arises is the one of *neighbor* and *neighborhood*. Two nodes are neighbors if there is a link among them, and the neighborhood of a node is the set of its neighbor nodes. Networks can be undirected, when links represent mutual interactions, or directed, like in the case of transport networks. In other words, in directed networks the proposition "i is contact with j" does not imply "j is contact with i". Furthermore, links can be binary relationships, or weighted by an intensity factor, modulating force-of-infection (weighted networks [128]). One of the advantages of the network approach is that a graph has a natural algebraic representation in terms of *adjacency matrix* [129, 130]. Given a network with N nodes labeled as $i, j = 1 \cdots N$, its adjacency matrix A has entry $A_{ij} = 1$ if i is connected to j, and zero otherwise. If the

network is undirected, then A is symmetric ($A = A^\dagger$); if the network is weighted, A_{ij} can assume values other than $0, 1$, encoding link weights. One of the most important graph-theoretical measures is *degree* of a node [80], i.e., the number of connections this node establishes with other nodes. For directed networks, we discriminate between incoming and outgoing degree, and for weighted networks we also introduce *strength* of a node, i.e. the sum of the weights of its links. Networks are often characterized in terms of their *degree distribution*, i.e., the statistical distribution of node degrees. When nodes establish links randomly, the resulting degree distribution is Poisson-like, with small dispersion around a mean value [80, 128]. On the other hand, networks with deeply non random connection patterns exhibit heterogeneous degree distributions, with large, sometimes diverging, variance. The most popular heterogeneous distribution in this context is the power-law $P(k) \sim k^{-\gamma}$ [131, 132], as many real networks are found to have degree distribution closely related to it. We summarize essential definitions of network theory in Appendix A.1. For a detailed introduction to networks, we refer to the books by Newman [80] and Barabási [83].

1.4. Epidemics on networks

In order to model the spread of a disease on a network, we choose the appropriate compartmental model. Then we infect a limited set of nodes (epidemic seeds). Assuming discrete time, at each time step all infectious nodes infect their neighbors with probability equal to disease transmissibility λ (see Figure 1.1). At the same time, in each node the disease progresses as required by the chosen model. For instance, in the case of SIR model, at each time step every infectious node recovers with probability μ.

Many features of real interaction networks have proven to dramatically impact the way diseases spread. The simplest, yet extremely crucial, among these features is the presence of hubs (large degree nodes). Heterogeneous degree distributions generally make the network more vulnerable to disease invasion and persistence, with respect to homogeneous mixing [76, 133–136]. As soon as the pathogen is able to infect the hubs, it gets access to a large part of the network, making its containment more difficult. Hubs also make the network more vulnerable to disease invasion (see [133] and Section 1.5.2), and more difficult to immunize. Strategies based on homogeneous mixing predicate that as soon as you immunize a critical fraction of the population, e.g. through vaccination, the disease is no longer able to turn epidemic. This phenomenon is called *herd*

immunity [137,138]. For homogeneous mixing this fraction is lower than 100%, and herd immunity can effectively be reached [28, 76]. In heterogeneous networks, instead, hubs have a higher spreading potential than the other nodes, and we will not be able to immunize the population, unless we target explicitly those hubs; random vaccination would require a coverage approaching 100% to reach herd immunity [139]. On the contrary, immunization strategies targeting preferentially hubs can effectively immunize the population [76,134]. Several methodologies for achieving this goal have been proposed, both assuming a global knowledge of the network [139], or a limited local knowledge [140].

Heterogeneous degree distribution is not the only feature that takes real social networks far from the homogeneous mixing assumption. Several other characteristics have been uncovered over the years, and have a great influence on how disease spread. In some networks, nodes with high degree have a higher than average probability of being in contact. These are called *assortative networks* [80]. If on the other hand they have a lower than average probability, they are *disassortative* [141]. Researchers have shown that assortative networks, as many social networks actually are, are more difficult to immunize [141, 142] than neutral or disassortative networks. The effect is opposite for the spread of cyber worms on the World Wide Web, which is disassortative [141]. Link weights have an impact on how diseases spread, too, as these change force-of-infection [143,144]. Weights can represent the duration of the contact [36,38], or, in case of transport networks, mobility fluxes [64, 71, 111]. Finally, clustering and community structures influence the spread of epidemics, by determining the conditions for global invasion, and the speed at which disease propagates [33].

1.5. The epidemic threshold

When a pathogen is introduced into a susceptible population, either it will cause an epidemic outbreak, or it will quickly go extinct. The outcome will depend both on disease features, and on the structure of contacts between hosts. Researchers in computational epidemiology aim at finding a way of discriminating these two conditions, in terms of the intrinsic transmissibility λ of the pathogen. Systems show the existence of a critical value λ_c of transmissibility, called *epidemic threshold* [30, 76], above which the disease is likely to turn epidemic. Conversely, when $\lambda < \lambda_c$, the outbreak will likely die out. Finding λ_c for a given system (i.e. population of agents) is crucial, as it allows both to predict the outcome of a potential pathogen introduction, and to assess the performance

of prevention strategies. The definition of epidemic threshold is intuitively clear and simple; its mathematical formulation, on the contrary, is not always straightforward, and require caution. We need do distinguish between disease models that allow for an endemic state, and models that do not. In the former category falls the SIS model and in general all models that do not confer permanent immunity: after an initial transient, the number of infectious individuals will saturate around a steady value. In the latter category we have the SIR model, and all models giving permanent immunity. Every infectious agent will eventually recover, until no more infections occur. Therefore, we will talk about SIS-like and SIR-like models [126].

For SIS-like models the epidemic threshold discriminates between the existence of an endemic state (above threshold), and the condition where the only stable steady state is the *disease-free state*, when there are no infectious agents (below threshold) [126, 145]. For SIR-like models, instead, the threshold is related to the final attack rate, i.e. the total fraction of agents that will have been infected during the whole course of the epidemic [126, 146]. SIR-like models above threshold will lead to a large final attack rate, while below threshold the attack rate will be negligible. Despite the epidemic threshold bearing different meanings in these two different families of models, conceptually it always discriminate between the epidemic and the extinction scenarios, and for this reason it represents a valid tool for assessing the vulnerability of a population to pathogen introduction. From the mathematical point of view, too, the epidemic threshold has a unified interpretation in terms of phase transition [147]. The two phases are clearly disease extinction and epidemic outbreak. By tuning model transmissibility (control parameter in the terminology of statistical mechanics) we pass from one phase to the other each time we cross the epidemic threshold. As we have seen, the measure that tells us in which phase we are (order parameter) is different for the two families of models. For SIS-like models it is the average fraction of infected agents once the disease reaches the endemic state, while for SIR-like models it is the final attack rate. The generic phase diagram is schematized in Figure 1.2. We observe the typical behavior of a phase transition: when one reduces transmissibility, the fraction of infected goes down, up to a tipping point where it becomes zero. Below this point it continues to be zero. Real systems, however, rarely show phase diagrams as "clean" as the one shown in Figure 1.2, as finite-size effects kick in, making the threshold more difficult to define and compute [148, 149]. Before delving into this point, we will review the computation of the threshold for continuous, deterministic models, for which everything is well defined.
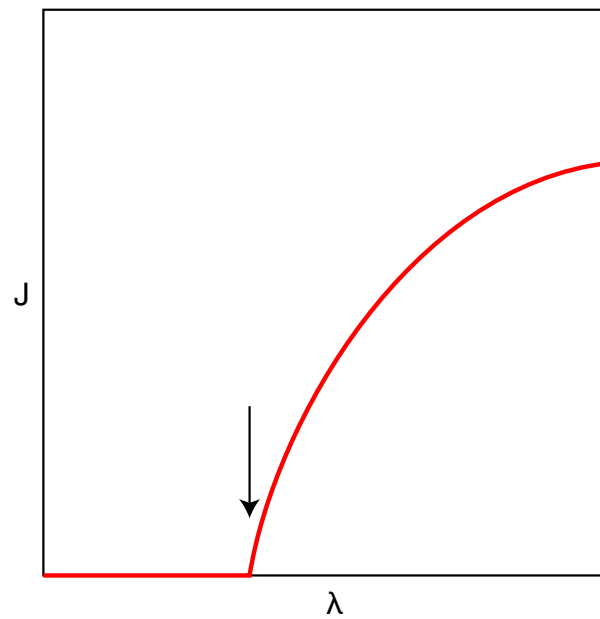
Figure 1.2.: **The epidemic threshold as a phase transition.** The value of order parameter J is plotted against disease transmissibility for an ideal setting. J corresponds to the final attack rate of SIR-like models, and to the average endemic prevalence for SIS-like models. The epidemic threshold is shown by the vertical arrow.

1.5.1. Homogeneous mixing: R_0

Equation 1.1 governs SIS dynamics in the framework of deterministic differential equations. Clearly, the disease-free state $I \equiv 0$ is a critical point, as $\dot{I} = 0$. Finding the epidemic threshold translates into finding when such critical point is stable (below threshold), as opposed to when a small number of infected ($I(0) > 0$) will bring the system far from $I = 0$. We can do this through linear stability analysis [150, 151], by keeping only terms that are linear in I, and remembering that population is constant ($S + I = N$):

$$\dot{I} \approx (k\lambda - \mu)\, I. \tag{1.3}$$

From this we directly see the critical value of the transmissibility: $\lambda_c = \mu/k$. We can generalize this result to more complex models, following the *next-generation matrix* approach [152]. Let us assume we have n_x non-infectious compartments $\{x_i,\ i = 1, \cdots n_x\}$, and n_y infectious compartments $\{y_\alpha,\ \alpha = 1, \cdots n_y\}$. We write the equations of a completely generic compartmental model, discriminating between infectious an non-infectious compartments:

$$\begin{cases} \dot{x}_i = f_i(x, y), \\ \dot{y}_\alpha = g_\alpha(x, y), \end{cases} \tag{1.4}$$

We concentrate on the equations for the infectious compartments, and note that we can split $g_\alpha(x, y)$ into a recovery part $g_\alpha^{(1)}(y)$, and an infection part $g_\alpha^{(2)}(x, y)$. Comparing this to SIS model (Equation 1.1), the only equation that matters is the second one, with $g^{(1)} = -\mu I$ and $g^{(2)} = \frac{k\lambda}{\mu} SI$. We now define two Jacobian matrices, one for each component of g:

$$\begin{cases} J_{\alpha\beta}^{(1)} = \left.\frac{\partial g_\alpha^{(1)}}{\partial y_\beta}\right|_{dfs}, \\ J_{\alpha\beta}^{(2)} = \left.\frac{\partial g_\alpha^{(2)}}{\partial y_\beta}\right|_{dfs}, \end{cases} \tag{1.5}$$

where both Jacobians are computed in the disease-free state (dfs), which is the equilibrium point that we wish to study, and where $y_\alpha = 0$. We now define the following key quantity:

$$R_0 = -\rho[J^{(2)}(J^{(1)})^{-1}], \tag{1.6}$$

where $\rho[X]$ is the **spectral radius** of matrix X, i.e., the largest among the absolute values of eigenvalues of X. R_0 is known in Epidemiology as **basic reproductive number** [28, 30, 153]. Matrix $J^{(2)}(J^{(1)})^{-1}$ is referred to as next-generation matrix [152]. R_0

is closely related to the concept of epidemic threshold, as it discriminates between the extinction phase $R_0 < 1$ and the outbreak phase $R_0 > 1$, as we show in Appendix A.4. Moreover, in some cases, R_0 has an interesting and concrete interpretation, that justifies its name. Let us consider the SIS and SIR models. They both have

$$R_0 = \frac{k\lambda}{\mu}. \tag{1.7}$$

In this case R_0 is made up of $1/\mu$, i.e. the average infectious period, λ, i.e. the rate of transmission, and k, i.e. the average number of contacts. Therefore, R_0 is **the average number of secondary cases generated by a primary case during its infectious period, in a fully susceptible population** [28, 30, 153]. In this sense, it is clear that an epidemic outbreak requires $R_0 > 1$.

The basic reproductive number has yet another interpretation: it is related to the exponential growth of the number of infected at the beginning of the disease. By solving Equation 1.3 we get $I(t) \approx I(0)e^{\mu(R_0-1)t}$, meaning that at the initial stage the disease grows exponentially at a rate tuned by $R_0 - 1$.

We have shown that the framework of deterministic differential equations allows a formal definition of epidemic threshold, in terms of basic reproductive number. We learn that for $R_0 > 1$ the disease will certainly turn epidemic, while when $R_0 < 1$ the growth is exponentially suppressed. We have already seen, however, that this kind of approach does not account for stochastic effects, and this qualitatively changes the definition itself of epidemic threshold. Let us examine this point through a Galton-Watson branching process [112–114]. Let I_n be the number of (infectious) agents in generation n. Also let $X = 0, 1, 2..$ be the number of secondary cases each case generates. $P(X)$ be the probability distribution of such variable, with probability generating function $g(s)$. By definition of basic reproductive number, $E[X] = R_0$. Let $g_n(s)$ be the probability generating function of I_n. Clearly, the following relation holds: $I_{n+1} = \sum_{j=1}^{I_n} X_j$, where X_j is the offspring of agent j in generation n. Given that X_j are identically distributed,

$$g_{n+1}(s) = g_n(g(s)). \tag{1.8}$$

From this relation we can compute the expectation values:

$$E[I_n] = g_n'(1) = g_{n-1}'(g(1))g'(1) = g_{n-1}'(1)g'(1) = E[I_{n-1}]R_0 = I_0 R_0^n. \tag{1.9}$$

Where I_0 is the initial number of infected. Analogously to differential equations, $R_0 = 1$ is a threshold value in the sense that discriminates the case when the average number of

infected grows or decays exponentially. These are however only expectation values, and do not account for stochastic fluctuations. These fluctuations influence the extinction probability, i.e. the probability that the disease will at some point stop due to lack of infectious hosts.

Let us call $q_n = Prob[I_n = 0]$, i.e. the probability of extinction by the $n-$th generation. Clearly $q_n = g_n(0)$. For simplicity, we assume just one initial infected: $I_0 = 1$. Then the following relation holds $g_n(g(s)) = g(g_n(s))$, and we can use it to get a recurrence relation fot q_n:

$$q_{n+1} = g(q_n). \tag{1.10}$$

We can find the limiting value q_∞ as the smallest positive root of

$$q = g(q). \tag{1.11}$$

$q = 1$ is always a solution of this equation. One can show that for $R_0 < 1$ this is also the smallest solution: $q_\infty(R_0 < 1) = 1$. When the disease is below threshold, the disease always goes extinct, as we already know from expectation values. In order to see what happens above threshold, we assume $X \sim \text{Poisson}(R_0)$. Then the above equation reads

$$q = e^{R_0(q-1)} \tag{1.12}$$

When $R_0 > 1$, this equation has a solution in $(0, 1)$: this means that the extinction probability is higher than zero even above threshold. As a result, if you run your epidemic long enough, it will always goes extinct. Therefore the epidemic threshold in stochastic models formally does not discriminate between extinction and no extinction, but between a phase where the disease goes extinct exponentially fast ($R_0 < 1$), and a regime where the time to extinction is long, and increases with R_0 (above threshold) [148, 154]. In practice the extinction time often becomes so long, that can be considered as infinite, recovering the meaning of the threshold found with differential equations. Formally, however, the active endemic state is always metastable, as fluctuations can always bring it to the absorbing state of no infected, which the system cannot then leave. These fluctuations become extremely suppressed only when R_0 is high, and/or the population is large.

We mention that the basic reproduction number, despite being a key quantity in epidemiology, remains a controversial concept, as several other methods computing it exist, and they do not always agree with one another [153]. We do not delve into this

subject, as it would take us far from the scope of this thesis. We use R_0 only as a tool for defining and interpreting the epidemic threshold ($R_0 = 1$).

The branching process is useful to introduce the stochasticity into the calculations, and see how this changes the meaning of the epidemic threshold. It does not, however, account for many features that we need to include. Since we assumed variable X is identically distributed for all agents, we are assuming there are no heterogeneities in the number of contacts each agent makes. In addition, since such distribution does not change in time, we assume each infectious agent has always access to a fully susceptible population.

In order to overcome these limitations, we will introduce the formalism of dynamical processes on networks.

1.5.2. Static and annealed networks

Different approaches have been developed for computing the epidemic threshold on networks, for the SIS and SIR models. Conceptually, they can be classified according to their way of taking into account the structure of the network [126, 136]:

quenched network The structure of the network is fixed, and expressed in terms of its adjacency matrix. The resulting epidemic threshold is characteristic of that particular network.

annealed network Only one (or more) statistical property of the network is relevant, i.e. its degree distribution. The threshold is then computed over the ensemble of all possible network configurations respecting that property.

Here we review two approaches based on annealed networks: one using differential equations [76, 133], and one using percolation theory [146]. We then review one approach based on quenched networks [155, 156], which is the most relevant to the development of our theory.

The first approach relying on annealed networks was developed by Romualdo Pastor-Satorras and Alessandro Vespignani [76, 133]. The threshold is computed as a function of a given degree distribution $P(k)$ of the network. Its original formulation requires the absence of degree-degree correlations, but can be generalized to include them [76, 142, 157]. Its core idea is to use the same differential equation approach that we have seen for homogeneous mixing (Equation 1.1), by splitting each compartment by degree. An

infectious node of degree k, and one of degree k' are no longer in the same compartment, but in I_k and $I_{k'}$, respectively. This approach is called *degree-block approximation*, or *degree-based mean field* [76,77,126], and assumes that all nodes with the same degree are statistically equivalent. Let S_k, I_k be the number of susceptible and infectious nodes in degree class k. Densitites are obtained by dividing S_k, I_k by the number of nodes with degree k: N_k.

$$\begin{cases} s_k = \frac{S_k}{N_k}, \\ i_k = \frac{I_k}{N_k}. \end{cases} \tag{1.13}$$

The average densities are obtained through the degree distribution $P(k)$: $s = \sum_k P(k)s_k, i = \sum_k P(k)i_k$. The system of equations governing SIS dynamics is

$$\frac{di_k}{dt} = \lambda k \sum_{k'} P(k'|k)i_{k'} - \mu i_k. \tag{1.14}$$

We study this system using linear stability around $i_k = 0$ (disease-free state). We wish to determine the conditions under which initial conditions around that point will lead to an exponentially growing outbreak. We rewrite the equations neglecting superlinear terms:

$$\frac{di_k}{dt} = \sum_{k'} \left[\lambda k P(k'|k) - \mu \delta_{kk'} \right] i_{k'} + \mathcal{O}(i_k^2) = \sum_{k'} C_{kk'} i_{k'} + \mathcal{O}(i_k^2), \tag{1.15}$$

in which we have defined matrix $C_{kk'} = \lambda k P(k'|k) - \mu \delta_{kk'}$. We assume now uncorrelated networks: $P(k'|k) = k'P(k')/\langle k \rangle$. For this specific case a close result can be derived. Treatment of the correlated case can be found in [76].

If we define two column matrices: $(C_1)_k = k$, and $(C_2)_k = kP(k)$, we can rewrite the uncorrelated C matrix as $C = \lambda C_1 C_2^\dagger / \langle k \rangle - \mu$. Therefore, C has rank equal to 1, and its principal eigenvector is C_1, with eigenvalue $\lambda \langle k^2 \rangle / \langle k \rangle - \mu$. By setting to zero the spectral radius of C: $\rho[C] = 0$, we recover the epidemic threshold:

$$\left(\frac{\lambda}{\mu} \right)_{critical} = \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{1.16}$$

Within the same framework it is possible to derive the threshold for the SIR model, too. Taking into account the fact that a node cannot re-infect the node it was infected by, the threshold becomes

$$\left(\frac{\lambda}{\mu} \right)_{critical} = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}. \tag{1.17}$$

These result are esepecially interesting in the case of heterogeneous degree distributions. In particular, when $P(k) \sim k^{-\gamma}$ with $2 < \gamma < 3$, the second moment of the

distribution diverges: $\langle k^2 \rangle \rightarrow \infty$ for $N \rightarrow \infty$. Vespignani and Pastor-Satorras [133] showed that in this case the epidemic threshold vanishes in the thermodynamic limit: $(\lambda/\mu)_{critical} \rightarrow 0$. For real systems, that are not infinite, this means that heterogeneous networks get indefinitely more vulnerable as they get bigger, because hubs get more and more connected, and can easily sustain an epidemic.

The second approach relying on annealed networks was developed by Mark Newman [146]. SIR dynamics is mapped onto a percolation problem [158], with link occupation probability equal to λ/μ. The threshold value of transmissibility is the minimum transmissibility value for which percolating cluster size grows as the size of the network. The threshold is derived in terms of probability generating functions (pgf) of two probability distributions: one is the degree distribution, and one is the probability of finding a node of a specific degree by following an edge picked at random. Newman shows that if $G_0(s)$ is the pgf of the former, then the pgf of the latter is $G_1(s) = G_0'(s)/\langle k \rangle$. Then he demonstrates that the epidemic threshold is

$$\left(\frac{\lambda}{\mu}\right)_{critical} = \frac{1}{G_1'(1)}. \tag{1.18}$$

This result converges to Equation 1.17 by noting that in case of uncorrelated networks $G_1'(1) = (\langle k^2 \rangle - \langle k \rangle)/\langle k \rangle$.

Finally, we review the approach based on quenched networks. By quenching we assume a fixed network, represented by its adjacency matrix A. Time is considered discrete, and at each time step a susceptible node can be infected by an infectious neighbor with probability λ. Analogously, it can recover with probability μ. Each node is assigned a binary stochastic variable X_i, corresponding to the infectious status ($X_i = 1$), and susceptible status ($X_i = 0$). Infection dynamic is translated into a discrete-time Markov process [159, 160] with 2^N possible states, corresponding to the possible infection configurations. The transition matrix among all these states cannot be written in general. For this reason, it is customary to neglect correlation among infectious statuses at fixed time [155, 156, 161]. Let us call $p_i(t) = \text{Prob}(X_i(t) = 1)$. According to this approach, called *quenched mean field*, it is possible to decompose the joint probabilities in terms of single node probabilities:

$$\text{Prob}(X_i(t) = 1 \text{ and } X_j(t) = 1 \text{ and } X_k(t) = 1 \cdots) = p_i(t)p_j(t)p_k(t)\cdots. \tag{1.19}$$

In this way the probability of a particular configuration occurring can be written in terms of single-node probabilities $p_i(t)$. The advantage is that for them an evolution

equation can be written down:

$$p_i(t+1) = 1 - [1 - (1 - \mu p_i(t))] \prod_j [1 - \lambda A_{ji} p_j(t)]. \tag{1.20}$$

This equation is the key to finding the threshold. For that, again we study its behavior around the equilibrium point $p_i = 0$. Analogously to what we did before, we wish to compute the minimum λ for which it is no longer a stable equilibrium, leading to an epidemic outbreak. In order to do that, we linearize the equation around $p_i = 0$:

$$p_i(t+1) = (1 - \mu)p_i(t) + \lambda \sum_j A_{ji} p_j(t) + \mathcal{O}(p_i^2). \tag{1.21}$$

This can be rewritten in matrix form, by interpreting $p_i(t)$ as an element of a size N vector $p(t)$:

$$p(t+1) = \left(1 - \mu + \lambda A^\dagger\right) p(t). \tag{1.22}$$

As a result, p moves away from 0 only when the largest eigenvalue (spectral radius) of $(1 - \mu + \lambda A)$ is larger than 1. From this we can find the threshold condition, as the minimum λ for which the following holds:

$$\left(\frac{\lambda}{\mu}\right)_{critical} = \frac{1}{\rho[A]}. \tag{1.23}$$

This result, derived for the SIS model, holds for the SIR model, too. In Section 3.3 we show that this is true for temporal networks as well.

In the quenched mean field approach, the epidemic threshold on a generic network of both the SIS and SIR model is completely determined by the spectral radius of the adjacency matrix.

2. Temporal networks

In the previous chapter we have seen that the structure of human interactions emerging from survey studies exhibited complex structures, that called for new mathematical tools: networks. Two main problems, however, affect these traditional ways of collecting contact data, and the reliability of the resulting contact networks. They are potentially prone to several sources of bias [162–164], which must be corrected for when using these data. Surveys make it difficult to get fine temporal resolution. Even when they give a good coverage of the contacts occurring within a time frame, they generally can provide neither the exact time at which they occurred, nor the order of their occurrence. As a result, traditionally, the temporal dimension wasn't taken into account in modeling the spread of a disease, simply because no fine enough data were available. In recent years this situation has dramatically changed. We now have ways to track contacts both in an almost unsupervised way, and with an extremely accurate time resolution. The former feature reduces the sources of bias, as individuals are no longer asked to report their contacts. The latter allows, for the first time, to delve into the dynamic evolution of contact networks. Developments in both hardware and software have driven this revolution. RIFD technology allows to record face-to-face proximity with high precision and accuracy [32, 35, 45] in several different contexts, like conferences [37, 39], museums [41], schools [33, 34, 38, 43]. We can track contacts among patients and health care workers in hospitals, through healthcare records [165–167] and again RFID technology [36, 40, 42, 44, 45]. Mobile communication [46–53] and transport [54–58] tracking shed new light on the dynamics of human mobility . Internet websites, forum and social networks can become a proxy for actual contacts, as it is the case for sexual encounters [59, 60]. In addition, contact and mobility data do not concern only our species: we can now study livestock mobility in Europe [61–66, 168], and proximity networks of farmed and wild animals [169–171]. This increased availability of high resolution data has prompted the need of new theoretical research in both network science and disease modeling. Researchers have shown that the interplay between network and disease dynamics

impacts the way epidemics break out, spread and persist in a population [39,87–92], making it necessary to find new theoretical tools that can handle these new time-resolved data. In this chapter we will define temporal networks, and describe the issues and challenges induced by adding the temporal dimension, in terms of their epidemiological relevance. We will introduce different representations of temporal networks, and how we model the spread of diseases on these new structures. We will then introduce some synthetic models of temporal networks, some null models, and some real contact datasets. All of them will be used in the following of our thesis.

2.1. Definition and basic properties

When contact patterns evolve in time, we need to go from static representations in terms of networks, to *temporal networks* [94, 95], i.e., networks whose links activate and deactivate in time. In terms of adjacency matrix (Appendix A.1), we go from a single matrix A for a static network, to a matrix which is function of time $A(t)$. Hence, while (static) networks are purely topologic objects, temporal networks have an embedded dynamic process driving link evolution in time. For this reason some call them *dynamic networks*. We stress that while the definition of temporal network is a straightforward extension of the definition of (static) network, the emerging properties of these new objects are conceptually and phenomenologically different, and cannot be in general recovered as a simple extension of what we know about static networks. The first conceptual problem arising from the temporal dimension is the definition of path and reachability. A path on a static network is a set of edges such that an edge ends where the next edge in the path begins. Paths matter a lot, as far as the spread of disease is concerned, because if there is a path going from node i to node j, it means that i can "influence" the state of j. To be more concrete, i has a possibility of infecting j if and only if a path from i to j exists. If the static graph is undirected, then every path connecting i to j will also connect j to i (see Figure 2.1A). This symmetry allow us to embed a metrics on the network, i.e., a rule for computing distances among nodes. We can do it only on static undirected networks because a metrics must be symmetric in its arguments: the distance from i to j must be the same as from j to i. In static directed networks this is no longer the case, as it may even occur that, while j is reachable from i - i.e. there exists a path from i to j - i is not reachable from j (see Figure 2.1B). As a result, directed networks are not metric spaces, and even reachability is not commutative.

It is, however, transitive: if $i \to j$, and $j \to k$, then $i \to k$ (see Figure 2.1B), where we use $\to$ in the sense *can reach*. Things get more complicated when we deal with temporal networks, for which none of the above properties holds. This happens because paths need to respect one more constraint: they need to be *time-respecting* [172]. Two links, connecting i and j and j and k, represent a path from i to k clearly only if $i - j$ activates before $j - k$ as in Figure 2.1C, otherwise, nothing in i will be able to reach k (Figure 2.1D). As a result, not only it is impossible to define a metric, but reachability is in general not transitive, as the chain $i \to j \to k$ implies $i \to k$ only if paths are arranged in a time-respecting way. These complex features of temporal networks matter a lot to us, because the spread of a disease on a network structure can be examined in terms of infection paths, which themselves must be time-respecting.

Time – the new ingredient of temporal networks – can be treated as a continuous or a discrete variable. If time is discrete, the temporal network is a sequence of snapshots which themselves can be seen as static networks. Assuming discrete time can be seen as an approximation, but it is both practically sound, and theoretically convenient. Indeed, most part of empirical datasets regarding temporal networks are intrinsically discrete. Take, for instance, livestock movements [61–66]; animal displacements are recorded on a daily basis, so the network naturally evolves in terms of daily snapshots. Analogously, proximity data recorder with RFID technology have an intrinsic time scale determined, which is around $20s$ to $30s$ [32, 35, 45], and below which it cannot detect contact with high enough precision and accuracy. As a result, discrete time becomes a natural choice in many empirical settings. For this reason, the theory we will develop in Chapters 3,5 assumes discrete time. There are, nonetheless, some contexts where this assumption does not hold. For this reason we will also extend our findings to the continuous-time case (Section 3.4).

2.2. Representations

Following [96], we divide the representations of temporal networks in *lossless* or *lossy*. Lossless representations carry all the information about the temporal network, and in practice we can identify them with the network itself. In lossy representation some information about the original temporal network is lost. Here we describe only the representations that are relevant for our thesis. For an exhaustive review we refer to [94, 96]. The most common lossy representation consists in projecting all the temporal
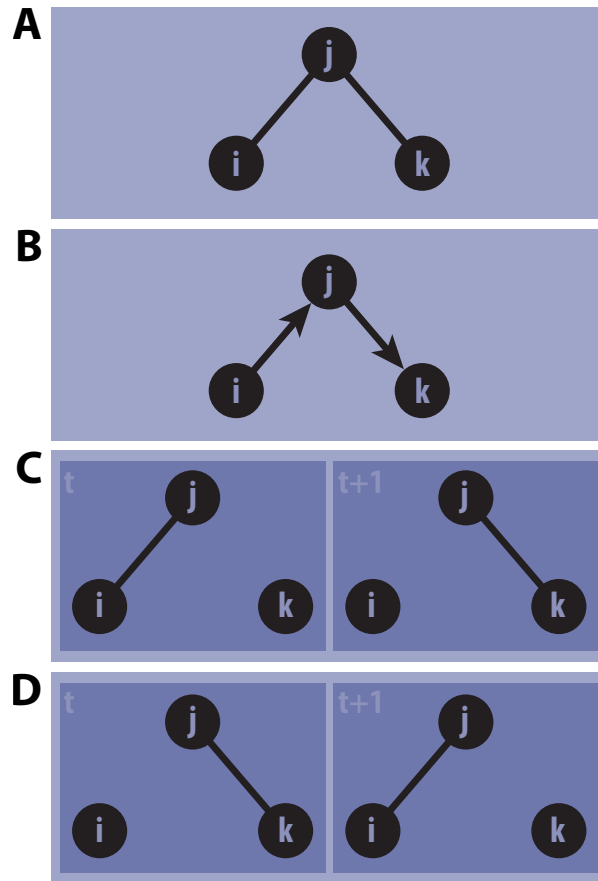
Figure 2.1.: **Path and reachability in static and temporal networks.** We consider a network of 3 nodes: i, j, k. **A** is a static undirected network. **B** is a static directed network. **C** and **D** are two temporal networks composed of two discrete-time snapshots. They comprise the same snapshots, but the order is reversed.

dynamics onto a static, aggregated, network. Here we consider two aggregation schemes, which we call HOM and HET, following the notation in [39]. HOM aggregation consists of building a single static network where two nodes are linked if they are in contact at least once in the temporal network. HET's topology is the same as HOM's, but HET links are weighted by the amount of time they are in contact in the temporal network. HET aggregation, despite losing all temporal correlations, can account for the fact that some ties are stronger than others, as they occur more frequently, or last longer. Researchers have compared the performance of these aggregation schemes in real settings [39], and provided theoretical evidence [173] of how heterogeneous tie strength can influence the way diseases spread. We study how well these aggregation models can be used to compute the epidemic threshold in Section 3.3 ([3]). Other studies have put forward more complex aggregation schemes [174]. They are designed to perform optimally in specific settings, where they can actually retain a large amount of meaningful structural information relevant to the spread of diseases. They are however out of the scope of this thesis.

Another static, lossy representation is the *accessibility graph* [94, 96, 175–178]. An accessibility graph is a directed graph in which a link going from i to j exists if there is a time-respecting path from i to j. It can be unweighted, or weighted by the number of time-respecting paths. Accessibility graph can be interpreted as a single static entity, encoding the whole temporal network, or as a growing object, containing only the time-respecting paths present up to a certain time. This is what Lentz and collaborators call *unfolding accessibility* graph [179]. Unfolding accessibility is of particular importance to us, as it is related to our methodology for computing the epidemic threshold on temporal networks, described in Chapter 3 and [2, 3]. In [179] the authors argue that the adjacency matrix accessibility graph, and in particular its time evolution, can be used to characterize the temporal network. In particular, they are able to recover the distribution of shortest path lengths, and provide a measure of how good the aggregated representation of the network is, in terms of how many paths present in the aggregated network are actually present in the temporal network. We report here the definition of the unfolding accessibility matrix as in [179].

$$\mathcal{P}_t = \prod_{h=1}^{T} \left(1 + A_t \right). \tag{2.1}$$

Time is assumed discrete, and A_t is the adjacency matrix of tth snapshot of the temporal network. More precisely, in [179] the authors are interested in the unweighted

version of the accessibility graph, so addition and multiplication should be intended as boolean operators: $+ \to OR$, $\cdot \to AND$. In [1, 3] (Chapter 3), we compare it to our *infection propagator* (Equation 3.2), and see that they share some common properties. In particular we will show that the accessibility matrix is a special case of our infection propagator, in the case of deterministic disease diffusion. The definition of accessibility matrix arises from a well known property of adjacency matrices, namely that $(A_t A_{t+1})_{ij}$ counts the number of paths that start at node i at time t and end in node j at time $t+1$. Moreover, the presence of the identity matrix in Equation 2.1 accounts for the fact that paths can contain waiting times. In other words, if we imagine a path as a route of a walker, without the identity term such walker would have to move from node to node at each time step. The identity terms allows it to stay still in one place indefinitely.

The need to add this identity term arises from an underlying conceptual difference between a path on a static network, and a path on a temporal network. A path on a static network is a purely geometric object, just as a curve joining two points on a plane. We have seen, on the other hand, that a temporal network is defined as a dynamic process. Hence, a temporal path from i to j can only be defined by propagating something from i to j according to some dynamical rules, like a random walker, or a spreading process. This apparently idle subtlety is the reason behind the representation we will use to compute the epidemic threshold on temporal networks, in which we merge network evolution and disease diffusion in a single dynamic process (Section 3.1).

We now turn to describing two lossless representations, which retain all the information of the temporal network. The first, most natural one, is the sequence of snapshots. Given a discrete-time temporal network of T time steps, we represent the links active at each time step as a static graph. The temporal network thus becomes a ordered sequence of graphs, or, algebraically, an ordered sequence of adjacency matrices $\{A_1, A_2, \cdots A_T\}$. The second is the multilayer representation, which retains a special importance for the development of our work.

A multilayer network is a network whose nodes can be grouped in different subsets (layers), with intra-layer and inter-layer connections having a different meaning [103]. Layers, too, usually carry a meaning beyond network topology; they can for instance represent different types of social interactions [180–182]. A temporal network formally becomes a multilayer network when we consider time snapshots as layers, with a copy of each node appearing in every layer. Given that multilayer networks can be handled using tensor algebra [102], we can define an *adjacency tensor* of a temporal network by

adding two "time indices" to the adjacency matrix, making it a rank 4 tensor. $A_{ts,ij}$ encodes the value of the link between i at time t and j at time s. Using the definitions in [102], we provide a rigorous definition of this tensor in Appendix A.2. The natural question that now arises is how to translate the links of the temporal network into the multilayer object. Following the nomenclature defined in [103], we can find in $A_{ts,ij}$ three different types of links, described in Figure 2.2. The most common way to proceed is to map the link of each snapshot onto the corresponding layer, so that each layer, with its intra-layer links, is a static representation of that snapshot. Then, in order to represent the temporal flow, each node is connected to its copy in the next layer through a inter-layer diagonal link. This representation has for instance been used to find communities in temporal networks [183–185]. The drawback, however, lies again in the fact that the multilayer network is a static geometric entity, while the temporal network is a dynamic process. For this reason, there is not clear way to differentiate – with different weights – inter-layer links representing temporal flow, to intra-layer links representing contacts. With no other information at hand, therefore, the intensity of inter-layer coupling must be arbitrarily fixed, or explored. In Chapter 3 we introduce a novel multilayer mapping that jointly represents both the temporal network and the spreading process. We will see that this merging removes this ambiguity in the coupling, as the disease itself fixes the value of all the coupling parameters.

2.3. Statistical properties

So far, we have dealt with topological features of temporal networks and their mathematical representations, in terms of their impact on epidemic spread. In many contexts we do not have, or we are not interested in, the exact structure of a network, but just the statistical distribution of some microscopic properties, and want to study how they influence diffusion. In the context of static networks, researchers have characterized the role of different centrality measures, like degree, betweenness, clustering. They have shown how different distributions of these quantities alter the epidemic outcome. Doing the same with temporal networks is not trivial, because one needs to find centrality measures that work for temporal network. A simple generalization of tools used in temporal network does not always lead to some meaningful result. Let us consider degree, for instance. Defined as the number of contacts, now degree changes in time. Then there is also the aggregated degree, which can be either the total number of connection made
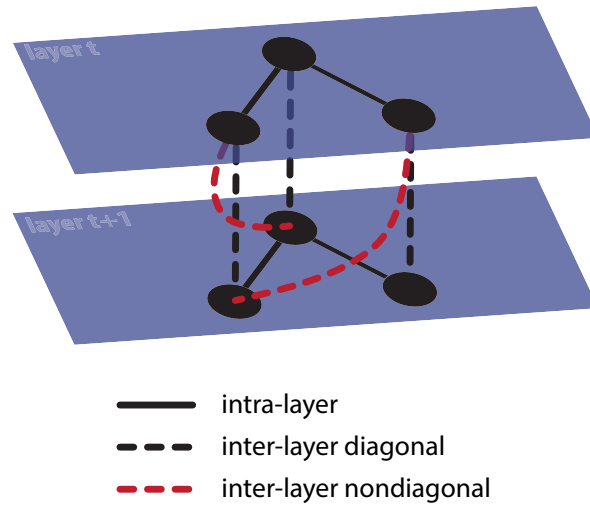
Figure 2.2.: **Temporal network as a multilayer object.** We represent a temporal network of 3 nodes and 2 time steps. We picture the three possible types of links. Intra-layer links (solid black) correspond to terms $A_{tt,ij}$ of the adjacency tensor. Inter-layer diagonal links (black dashed) correspond to $A_{ts,ii}$ terms. Inter-layer non-diagonal (red dashed) correspond to terms $A_{ts,ij}$, with both $t \neq s$ and $i \neq j$.

(HET), or the number of nodes met at least once (HOM). These measures, however, have proven to be ineffective in some contexts, as they say nothing about the epidemic risk. This effect visible, for example, in livestock networks, where animal holdings change their number of contacts over time so dramatically, that hubs can become poorly connected nodes in a short period, and vice versa, with no easily predictable patterns. That makes their instantaneous degree difficult to predict, and their aggregated degree not representative of their behavior [64]. As a result, we clearly need new centrality measures that account for the timely behavior, and can be used as simple risk factors. In the context of information diffusion, Miritello et al. [186] define *social strategy* using both instantaneous degree and aggregated degree [186]. Let us consider a discrete-time temporal network of T snapshots. In [186] they fix a time window of δ, and a time t. We aggregate the network in the interval $[t - \delta, t]$, following the HET scheme. For every node i we consider k_i and s_i as its degree and strength, respectively (see Appendix A.1 for definitions). Social strategy γ_i of node i at time t, and with time window δ, is then defined as

$$\gamma_i(t) = \frac{k_i}{s_i}. \tag{2.2}$$

This is the definition we will use in this theses, and it is analogous to the one introduced in [186], except for a normalizing factor δ. $\gamma \to 0$ means that the node makes almost all its contacts always with the same set of nodes, while the opposite regime, $\gamma \to 1$ characterizes a node which changes its neighbors often. Social strategy is thus able to discriminate between a memory-driven behavior, and an exploratory one. We will use this measure in Section 3.3 and [3], to measure the social behavior of students in a high school (Section 2.4).

In Section 5.3 and [1], when dealing with two datasets, Italian cattle trade movements (Chapter 4) and human sexual contacts (Section 2.4), we introduce a novel measure: *loyalty*. Loyalty, like social strategy, aims at assessing node memory. However, while social strategy is averaged over a time window, loyalty quantifies the behavior of a node from one snapshot to one that immediately follows. We demand this because, as we will see, we aim at providing a risk prediction methodology that requires the least amount of past contact data, as we usually do not have a long time window to exploit. Specifically, loyalty measures the fraction of preserved neighbors of a node for a pair of two consecutive network configurations in time. Let us consider snapshots t and $t + 1$, and let $\mathcal{V}_i^t$ be the set of neighbors of node i in snapshot t. Then loyalty of node i from

time t to time $t + 1$ ($\theta_i^{t,t+1}$) is defined as the Jaccard index between $\mathcal{V}_i^t$ and $\mathcal{V}_i^{t+1}$:

$$\theta_i^{t,t+1} = \frac{\left|\mathcal{V}_i^t \cap \mathcal{V}_i^{t+1}\right|}{\left|\mathcal{V}_i^t \cup \mathcal{V}_i^{t+1}\right|} \tag{2.3}$$

Loyalty takes values in $[0, 1]$. $\theta = 0$ means no neighbors are retained, while $\theta = 1$ means that the node keeps exactly the same neighbors from t to $t + 1$. In the case of cattle trade network, which is directed, we choose to consider in-neighbors of i, i.e., nodes from which i buys cattle from. In this sense we can interpret "loyalty" as the willingness of farm i to keep buying cattle always from the same providers. In addition to Jaccard index, loyalty can also be expressed in the adjacency matrix formalism, as follows:

$$\theta_i^{t,t+1} = \frac{\sum_j (A_t A_{t+1}^\dagger)_{ji}}{\sum_j \left[A_{t,ji} + A_{t+1,ji} - (A_t A_{t+1}^\dagger)_{ji}\right]}. \tag{2.4}$$

2.4. Datasets

We introduce here some datasets recording high-resolution host interactions, which are naturally modeled as temporal networks, and which we will use in this thesis. They cover different types of interactions, all extremely relevant from the epidemiological perspective. We present two datasets recording human proximity interactions, which are a potential route for the spread of respiratory viruses like flu [38], or bacterial infections that can be transmitted through direct contacts, like Staphilococcus Aureus [45, 187]. We then describe a network of sexual contacts, through which we can study study the spread of Sexually Transmitted Diseases (STDs), like syphilis (Section 5.2). We will not limit ourself to the human species. We will present also the network of cattle trade displacements in three European countries, which is extremely important for the spread of diseases among farmed animals. Given the novelty and length of this analysis, we will devote an entire chapter to it (Chapter 4).

Face-to-face interactions make use of Radio Frequency Identification (RFID) technology [32]. Individuals wear RFID tags on their chest. These tags can communicate among themselves and with antennas by exchanging wave packets in the microwave band. Whenever two tags are in contact, they signal their proximity to the closes antenna, which records the identity of the interacting tags and the time. Wave packet intensity is purposely low, so that tags record mutual proximity only when they are closer than approximately $1m$. Moreover, water in human body is able to shield wave

packets completely, so that only face-to-face proximity is detected. This technology works best in closed environments, like schools [33, 38, 43], conferences [37, 39], offices, museums [41], hospital wards [36, 42, 45], where rooms can be equipped with receiver antennas. Time evolution of these networks is intrinsically discrete, as a certain number of wave packets must be received in order to confirm the statistical significance of the contact, with respect to noise. Time resolution is however generally high, around $20s$.

In our works we use

HT09 Face-to-face interactions during a scientific conference, recorded by the Sociopatterns[1] group [37] (see Figure 2.3**A**,**B**);

SCHOOL Face-to-face interactions in a high school, recorded by Salathé and collaborators [33] (see Figure 2.3**C**,**D**);

SEX Sexual encounters between female prostitutes and their male clients, recorded by Rocha and collaborators though an escort rating web forum [60] (see Figure 2.4).

Table 2.1 provides a basic description of these datasets, and Figure 2.3,2.4 show the temporal behavior of these networks, in terms of their *activity timelines* (number of active contacts as function of time) and inter-activation time distributions.

dataset id	nodes	duration	highest resolution	ref
HT09	113	2.5 days	$20s$	[37]
SCHOOL	787	7 hours	$20s$	[33]
SEX	12500	2.5 years	1 day	[60]

Table 2.1.: **Human contact networks used.**
We list some basic features of the datasets we use. Sometimes they differ from the numbers in references, as we discard initial and final transients.

2.5. Synthetic models of temporal networks

The shape of real networks is a result of several driving forces, whose effects are difficult to disentangle. As a result, these networks are generally highly correlated in space and time. Gauging the impact of a specific property on disease diffusion is, therefore, hard.
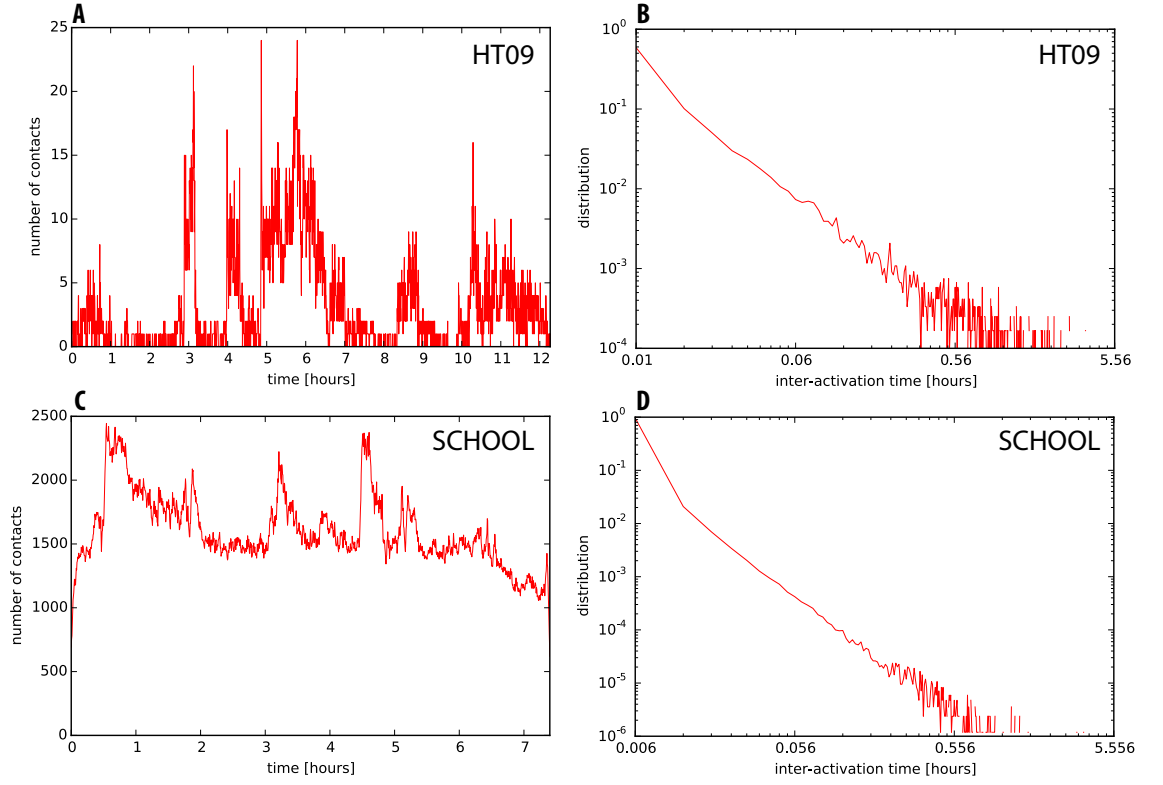
---

[1]www.sociopatterns.org

Figure 2.3.: **Time evolution of** HT09 **and** SCHOOL **networks.** For both networks we show activity timeline (HT09: **A**, HT09: **C**) and the distribution of node inter-activation time (HT09: **B**, HT09: **D**). We show the whole SCHOOL dataset, and a full day of conference for HT09. In both we keep the finest possible resolution, i.e., $20s$.
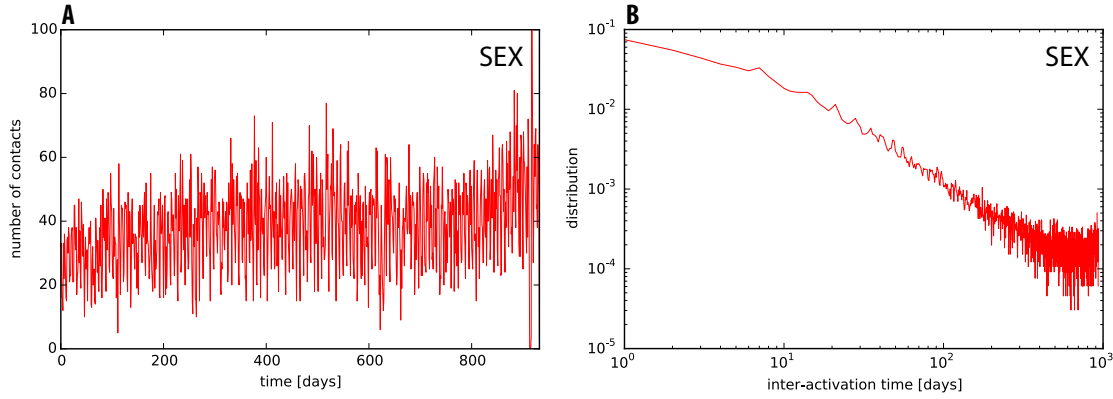
Figure 2.4.: **Time evolution of** SEX **network.** We show activity timeline (**A**) and the distribution of node inter-activation time (**B**).

To overcome this issue, we make use of synthetic generative network models [91, 100, 174, 188–192]. A model is designed to exhibit just the features we wish to investigate. Furthermore, we can tune the intensity and interplay between different characteristics, and assess their impact on disease diffusion in all possible regimes. We will use four different models, which are now introduced. For all of them, we assume discrete time.

ER **and** ER-T**: Erdős-Rényi sequences**

A random graph – or Erdős-Rényi graph – with N nodes and M edges is a random instance of the set of all possible graphs with those numbers of nodes and edges [193]. In practice, we fix a probability p, and connect each possible pair of nodes with that probability. The two definitions converge for large N, provided $M = pN(N-1)/2$. The degree distribution of an Erdős-Rényi graph is thus a binomial with $N-1$ as number of trials, and p as success probability: $k \sim \text{Binomial}(N-1, p)$. In particular, if we are interested in a large network with fixed average degree $\langle k \rangle = p(N-1)$, then $k \sim \text{Poisson}(k)$. An Erdős-Rényi graph is therefore homogeneous, in the sense that it does not exhibit large degree fluctuations: $\sqrt{\text{Var}(k)}/k = 1/\sqrt{k}$. An Erdős-Rényi graph, being completely uncorrelated and homogeneous, is a perfect null model for real networks, which often exhibit heterogeneous degree distributions [131] and correlations [76]. We extend this model to the temporal case, by building a sequence of uncorrelated Erdős-Rényi networks (ER), all with the same edge creation probability. In addition to the absence of

within-snapshot correlations, ER has no temporal correlations as well. Moreover, network evolution is Markovian: each node (or edge) has no memory of its past behavior. The number of time steps Δt occurring between two consecutive activations of the same edge is thus geometrically distributed: $\Delta t \sim p(1-p)^{\Delta t-1}$, which is the discrete equivalent of the exponential distribution, typical feature of memoryless dynamics. Again many real systems exhibit inter-event time distributions which deviate largely from the exponential distribution, with power law-like behavior, and large variance [64, 90, 91, 94, 126]. In this sense, many real networks exhibit a heterogeneous behavior for inter-contact times, too, as opposed of the more homogeneous behavior (smaller variance) of ER.

ER-T is a variation of ER, in which again microscopic correlations are present neither in topology, nor in time. This time, however, we allow for a time-varying activation pattern: the number of links in each snapshots – equivalently, its average degree – exhibits externally driven fluctuations. We implement this by letting the edge creation probability depend on time: $p(t)$. Let us define the *activity timeline* of a network the number of active contacts in each time step. Analogously we could use average degree, or any measure that gives a macroscopic, coarse-grained information of the level of activity in time. Then ER-T is useful to enforce a non-trivial activity timeline and to study, for instance, the impact of periodic, seasonal-like variations in the global activity of the network, decoupled from microscopic memory effect and correlations. We can characterize ER-T as non-autonomous, as its parameters depend on time. We will use ER-T in Section 3.4.

ACTIVITY: **Activity-driven model**

This model was put forward by N. Perra and collaborators [100]. Its main goal is to provide a simple mechanistic explanation of the evolution in time of a realistic network, that could lead to the observed heterogeneous degree distributions of the aggregated representation. They assign each node an activity potential a_i, sampled from a given distribution $f(a)$. At each discrete time steps, nodes become active each with probability a_i. Active nodes establish each m connections with other nodes, selected randomly among both active and inactive ones. Degree distribution in each snapshot is homogeneous, as it is binomially distributed for both active and inactive nodes. As links are renewed at every snapshot, the process is memoryless, and the inter-activation time distribution is geometric, just as in ET. And indeed ET (with edge-creation probabil-

ity p) can be seen as a special case of ACTIVITY, with constant activity $a_i = a$, and $p = 2ma(1-a)/N$. However, ACTIVITY's interest lies in the opposite regime, i.e., when a distribution is broad. They show the aggregated degree distribution over T time steps has the same shape of the activity distribution: $P(k) \sim f(k/Tm)$, in the limit of small k/N and k/T. In [100] they set $f(a) \sim a^{-\gamma}$, and $a \in [\epsilon, 1]$. With this choice, they manage to recover a power law shaped aggregated degree distribution. While the topology of each snapshot is extremely simple, the heterogeneous behavior in time evolution is thus able to produce the rich phenomenology we observe in real network. Another advantage of this model is that it is analytically treatable, and in particular its epidemic threshold can be computed using heterogeneous mean field (see Chap. 3). A second version of the activity driven model also exists, and integrates non-Markov behavior in time evolution, namely nodes retain memory of their past dynamic, and when they establish links, they are likelier to choose past neighbors [173].

BURSTY**: Heterogeneous inter-activation time**

Both ER and ACTIVITY share a homogeneous inter-activation time distribution. However, we've seen how many real systems exhibit heterogeneous ones. This is the rationale behind the BURSTY model, defined by L. Rocha and collaborators in [91]. They first fix a distribution $g(t) \sim t^{-\alpha}e^{-\beta t}$, which is a power law (i.e., heterogeneous), with an exponential cutoff. At each time step, each node computes the time passed since last time it had been active: Δt, then activates with probability $g(\Delta t)$. In this way inter-activation time is forced to be heterogeneous.The overall level of heterogeneity can be tuned using the parameters: $\alpha = 0$ is completely homogeneous and Markov-like, while $\beta = 0$ is fully heterogeneous. At each time step active nodes connect among themselves, with just one link each. Consequently, in each snapshot, active nodes have all degree equal to 1, but the model can be straightforwardly generalized to an arbitrary degree k.

Tab. 2.2 summarizes the main features of the models we have defined.

2.6. Null models for temporal networks

We have seen that, unlike real networks, models allow us analyze the impact of each specific temporal property, disentangling it from the others. They are, however not the only way. Indeed we may decide to adopt the opposite course of action, start from the real temporal network, and destroy in turn some of its features, obtaining several *null*

model	instantaneous degree (k)	aggregated degree (k)	inter-activation time (Δt)	au-tonomous
ER	hom ($k \sim$ Poisson)	hom ($k \sim$ Poisson)	hom ($\Delta t \sim$ Geometric)	yes
ER-T	hom	hom	hom	no
AC-TIV-ITY	hom	het ($k \sim k^{-\gamma}$)	hom ($\Delta t \sim$ Geometric)	yes
BURSTY	fixed	hom	het ($\Delta t \sim \Delta t^{-\alpha} e^{-\beta \Delta t}$)	yes

Table 2.2.: **Models of temporal networks, and their basic features.**

models. Conceptually, while a model allows you to control exactly which features you put in, a null model of a real network lets you choose what features you take out, leaving all the residual properties, which you may or may not know. We describe here some null models, which serve different purposes.

Firstly, RESHUFFLE is a randomization of snapshot order. Secondly, let us define a timestamped contact as a triplet (i, j, t), i.e., a link between i and j occurring at time step t. RECONFIGURE consists in a random reassignment of contact timestamps: Two contacts (i, j, t), (k, l, s) are randomly selected, and their timestamp switched: (i, j, s), (k, l, t). At the macroscopic level, both null models preserve the aggregated network. RESHUFFLE clearly destroys the activity timeline, breaking all possible seasonal or circadian patterns. RECONFIGURE, on the other hand, preserves it. Microscopically, RESHUFFLE preserve link correlations inside each snapshot, while RECONFIGURE does not. They both destroy all time correlations and self-correlations in link activation: memory is lost, bursty inter-activation time is lost, time-respecting paths and temporal motifs (correlation patterns in link activations [64, 194]) are broken. Finally, in ANONYMIZE we randomize the identity of the nodes, independently inside each snapshot. ANONYMIZE preserves only the activity timeline, while destroying anything else, including aggregated network. Tab. 2.3 summarizes the features of all these null models.

model	activity timeline	aggregated network	snapshot topology	microscopic time correlations
RESHUF-FLE	destroy	preserve	preserve	destroy
RECON-FIGURE	preserve	preserve	destroy	destroy
ANONYMIZE	preserve	destroy	destroy	destroy

Table 2.3.: **Null models for real temporal networks: What they destroy and what they preserve.**

2.7. Epidemics on temporal networks

If we represent the temporal network in terms of a sequence of snapshots, the way of simulating the spread of a disease is similar to the static case (Section 1.4). Recovery and the other spontaneous processes which do not depend on the network are exactly the same as in the static case. Then, at each time step we let the disease spread along the contacts that are active in the corresponding snapshot. The only issue is when the duration of the temporal network is shorter than the epidemic wave we wish to study. In this case we have to make up for the missing contact data, by re-using contacts in past configurations. The most common way to do it is to impose periodic boundary conditions: when we reach the last snapshot of the network, we plug in the first one and repeat the network as long as we need to. Formally, this clearly breaks causality, and that may induce unwanted effects. Practically, however, if the data collection period is long enough, spurious effects due to causality violations are suppressed enough, and periodic conditions can be safely enforced. In particular, in Section 3.2 and [1] we show that periodic boundary conditions do not impact the epidemic threshold in several synthetic models and real networks.

In [39] the authors explore other ways to reuse past data beyond periodic conditions, by reshuffling, for instance, node identities both randomly, and respecting some node-activity correlations.

3. The epidemic threshold on temporal networks

As we learned in Chapters 2,4, many contact structures relevant for the spread of several diseases evolve in time. Human face-to-face interactions, responsible for the spread of airborne diseases, or sexual interactions, relevant for the diffusion of STDs, all result in highly dynamic contact networks (see Section 2.4). Livestock transport networks, which are responsible for the spread of many animal diseases, exhibit a highly dynamic nature, too, as we have seen in Chapter 4. The interaction between network evolution and disease diffusion is known to influence the dynamics and outcome of the epidemic process (see [37, 59, 87–92, 97–100] and Chapter 2). Specifically, such interaction alters the epidemic threshold (Section 1.5), affecting the conditions that lead to the wide-spreading regime. For this reason, being able to compute the epidemic threshold, while accounting for network dynamics, is a crucial step towards accurately assessing the vulnerability of the system to a particular pathogen. Up to now, researchers have analytically computed the epidemic threshold in the two limiting scenarios described in Section 1.5.2; quenched and annealed. These are accurate in the regime of time scale separation, i.e., when network evolution and disease diffusion occur at very different paces, with the network evolving much slower than disease (quenched) or much faster (annealed). When the two time scales are comparable, and thus fully coupled, researchers have so far resorted to numerical simulations, or provided analytical calculations only for specific cases [92, 97–101]. In this chapter we introduce a particular multilayer mapping that allows us to analytically compute the epidemic threshold of a generic temporal network. In the first part (Sections 3.1,3.2) we introduce the methodology, we test its accuracy on different synthetic networks and real datasets, and also explore how the overall time of observation of the temporal network impacts the accuracy of the threshold. We then devote the following of the chapter to extend it to general settings and different disease models. More in detail, in Section 3.3 we generalize our computation of the threshold

to weighted and directed networks, in order to account non-symmetric contact patterns and heterogeneous force-of-infection. We also go beyond the SIS model, allowing for temporary and permanent immunity (SIRS and SIR models).

Our multilayer approach intrinsically assumes discrete time. Given, however, that in some contexts temporal networks are modeled through continuous-time dynamics [87,90], in Section 3.4 we derive an equation for computing the threshold in continuous time, and explicitly solve it in a particular case.

Finally, in Section 3.5, we study the impact of disease latency period on the epidemic threshold. We realize that, unlike recovery, latency cannot be included into our computation as a simple extension. We thus develop a novel methodology that, again through multilayer tools, is able to compute the epidemic threshold of the SEIS model on a generic temporal network. We find that latency impacts the value of the epidemic threshold in many contexts, unlike what happens in the static case, where the threshold depends only on the infectious period. We then explore what are the features of network dynamics that, through coupling to latency period, alter the epidemic threshold.

3.1. A novel multilayer mapping of network evolution and disease dynamics

Here we present a novel multilayer mapping of the temporal network, which we will use to compute the epidemic threshold. This mapping is contained in the articles [1, 3] which make up the following sections. Nonetheless, given its crucial importance in our threshold computation methodology, we devote this section to its description.

Let us consider a discrete-time, undirected, unweighted temporal network, of N nodes and T time steps. We will extend the methodology to weighted, directed in Section 3.3, and take care of continuous time in Section 3.4. Let $\{A_1, A_2, \cdots A_T\}$ be the sequence of adjacency matrices representing the snapshots. We now define a multilayer structure composed of T layers, one for each snapshot, each containing a copy of all the N nodes. Hence, each node of the multilayer is identified by a pair (i, t), with $i = 1, \cdots, N$ and $t = 1, \cdots, T$. We define links among layers as follows:

- each node in layer t is connected to its own image in layer $t + 1$ via a directed link (orange arrows in Figure 3.1);

- a link between i and j at time t ($A_{t,ij} = 1$) translates into a pair of directed links,

one going from i at in layer t to j in layer $t+1$, and one going from j at in layer t to i in layer $t+1$ (black arrows in Figure 3.1);

- we enforce periodic boundary conditions: layer T is connected to layer 1 with the above rules, assuming the equivalence $T+1 \equiv 1$.

We now call $\mathbf{A}$ the adjacency tensor of the multilayer network, as defined in Section 2.2. The only non-zero terms in $\mathbf{A}$ are $\mathbf{A}_{t,t+1,ii} = 1$ and $\mathbf{A}_{t,t+1,ij} = A_{t,ij}$. We stress that in our multilayer there are no intra-layer links, resulting in a multipartite structure. In Figure 3.1 we schematically show the structure of our multilayer structure. The rationale

Figure 3.1.: **Multilayer representation of the temporal network.** We show the rules to build our multilayer representation of the temporal network. We focus on 3 nodes and 2 time steps.

behind this representation lies in the spreading process that we wish to couple to it. Let us consider a SIS spreading dynamics, with transmissibility λ and recovery probability μ. We suppose now that node i is infectious at time t. If nodes i and j are connected at time t, there is a chance (λ) that i will infect j, resulting in an infectious j at time $t+1$. This explains why link $i-j$ at time t is represented through non-diagonal links going from layer t to $t+1$ (black arrows in in Figure 3.1). Moreover, going from t to

$t+1$, node i may either recover or remain infectious. We interpret the probability that i does not recover, as the probability it does not transmit the disease to its future image at $t+1$. We implement that through the diagonal inter-layer links (orange arrows in in Figure 3.1). Given that recovery occurs with probability μ, the probability that infection is passed from i at time t, to i at time $t+1$ is $1-\mu$. In addition we stress that the absence of undirected links and intra-layer links in the multilayer structure guarantees that causality in temporal evolution is preserved.

The mapping we have defined thus naturally couples network dynamics to disease diffusion, which fixes the intensity of the different types of links. Non-diagonal links will have a weight λ as they transmit the pathogen among different nodes, while diagonal links will be weighted by $1-\mu$, encoding recovery dynamics.

Instead of identifying each node of the multilayer with (i,t), we can consider the whole network as composed of NT distinct nodes, flattening out the layer structure. Nodes are now identified by $\alpha = 1, \cdots, NT$, with $(i,t) \to \alpha = Nt + i$. Algebraically, we are exploiting the fact that the tensor space $\mathbb{R}^N \otimes \mathbb{R}^T$ in which the multilayer lives is isomorphic to $\mathbb{R}^{NT}$. Thanks to this, instead of dealing with a tensor, we can now switch to the supra-adjacency matrix formalism [102,195,196], by writing the adjacency matrix of the NT graph. Such matrix will be of size $NT \times NT$, and composed of T^2 blocks of size $N \times N$, representing the original layers. In addition, given that we already have added disease dynamics, we weigh links accordingly. The resulting matrix is

$$
M = \begin{pmatrix}
0 & 1-\mu+\lambda A_1 & 0 & \cdots & 0 \\
0 & 0 & 1-\mu+\lambda A_2 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 1-\mu+\lambda A_{T-1} \\
1-\mu+\lambda A_T & 0 & 0 & \cdots & 0
\end{pmatrix}, \quad (3.1)
$$

in which each term indicates a block.

In the next section we show that this matrix accurately models the diffusion of a SIS process on the temporal network $\{A_1, A_2, \cdots A_T\}$, and provides the key to compute the epidemic threshold.

3.2. Article: Analytical computation of the epidemic threshold on temporal networks

In the following article [1], using the multilayer mapping we have just introduced, we compute the epidemic threshold of the SIS model on a generic temporal network. We find that the threshold can be found using the following matrix, which we dub *infection propagator* [3]:

$$P = \prod_{h=1}^{T} \left[1 - \mu + \lambda A_h \right]. \tag{3.2}$$

In particular, the threshold is the value of the transmission parameter λ for which the spectral radius[1] of P is equal to 1: $\rho[P(\lambda_c, \mu)] = 1$. In the article we also test the performance of our calculation against numerical simulations in different network models, and real datasets. Moreover, by studying how the epidemic threshold varies as a function of T, we show that through our methodology we can compute the optimal data collection time interval for an experimental contact tracing setting.

---

[1]The spectral radius $\rho[X]$ of matrix X is the maximum of the absolute values of its eigenvalues. See also Section 1.5.1.

BEGINNING

Valdano E, Ferreri L, Poletto C, Colizza V

Analytical Computation of the Epidemic Threshold on Temporal Networks

Analytical Computation of the Epidemic Threshold on Temporal Networks

Eugenio Valdano,[1,2] Luca Ferreri,[3] Chiara Poletto,[1,2] and Vittoria Colizza[1,2,4,*]

[1]*INSERM, UMR-S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique,*
F-75013 Paris, France
[2]*Sorbonne Universités, UPMC Univ Paris 06, UMR-S 1136,*
Institut Pierre Louis d'Epidémiologie et de Santé Publique,
F-75013 Paris, France
[3]*Dipartimento di Scienze Veterinarie, Università degli Studi di Torino, Grugliasco (TO) 10095, Italy*
[4]*ISI Foundation, Torino 10126, Italy*
(Received 18 August 2014; published 8 April 2015)

The time variation of contacts in a networked system may fundamentally alter the properties of spreading processes and affect the condition for large-scale propagation, as encoded in the epidemic threshold. Despite the great interest in the problem for the physics, applied mathematics, computer science, and epidemiology communities, a full theoretical understanding is still missing and currently limited to the cases where the time-scale separation holds between spreading and network dynamics or to specific temporal network models. We consider a Markov chain description of the susceptible-infectious-susceptible process on an arbitrary temporal network. By adopting a multilayer perspective, we develop a general analytical derivation of the epidemic threshold in terms of the spectral radius of a matrix that encodes both network structure and disease dynamics. The accuracy of the approach is confirmed on a set of temporal models and empirical networks and against numerical results. In addition, we explore how the threshold changes when varying the overall time of observation of the temporal network, so as to provide insights on the optimal time window for data collection of empirical temporal networked systems. Our framework is of both fundamental and practical interest, as it offers novel understanding of the interplay between temporal networks and spreading dynamics.

Subject Areas: Complex Systems,
Interdisciplinary Physics,
Statistical Physics

I. INTRODUCTION

A wide range of physical, social, and biological phenomena can be expressed in terms of spreading processes on interconnected substrates. Notable examples include the spread of directly transmitted infectious diseases through host-to-host contacts [1], the spatial propagation of epidemics driven by the hosts' mobility network [1–3], the spread of cyber viruses along computer connections [4], or the diffusion of ideas mediated by social interactions [5,6]. These phenomena are the result of a complex interplay between the properties of the spreading dynamics and the network's structural and temporal features, hindering their full understanding.

A fundamental issue characterizing spreading processes is the identification of the critical condition for the wide-spreading regime, encoded in the epidemic threshold parameter. This issue is of critical importance for epidemic containment [1], as well as for control of the diffusion of information [7] and cyber viruses [4]. Extensive studies have characterized this parameter in the time-scale separation approximation, i.e., when the time scales of the spreading process and network evolution strongly differ. This includes the two limiting regimes, quenched and annealed [4,8–15]. In the first case, the network is regarded as static, as it evolves on much slower time scales than the ones characterizing the spreading process. The epidemic threshold in this case is computed from the adjacency matrix describing the network connectivity pattern [8,9]. In the second case, the underlying network evolves so rapidly with respect to the dynamical process that only its time-averaged properties are relevant to the spreading dynamics. Approaches like the heterogeneous mean field [4], the generating function [11], and percolation theory [10] provide, in this regime, estimates of the threshold in the infinite size limit.

[*]Corresponding author.
vittoria.colizza@inserm.fr
http://www.epicx-lab.com

Recently, the extensive empirical characterization of social interactions at different scales and settings [16–22] has shown that networks often display non-Poissonian and non-Markovian temporal evolutions unfolding at time scales similar to the ones of many spreading processes of interest, stressing the need for novel theoretical tools able to overcome current limitations. Much research has focused on spreading processes occurring on time-varying networks [16,17,21,23–31], modeled either as a discrete-time sequence of networks [16,28] or as continuous-time dynamics of links [17,26]; however, so far, only a few studies have provided an analytical calculation of the epidemic threshold in specific cases [24,27–32]. These are all based on models for time-varying networks integrating the microscopical laws governing the network evolution, under context-specific assumptions. An analytical framework for the computation of the epidemic threshold for an arbitrary time-varying network is still missing. To fill such a gap, we present here a novel approach that, by reinterpreting the tensor formalism of multilayer networks [33,34], extends the Markov chain approach adopted for static networks [8,9] to their temporal counterpart. The approach is applied to discrete time-varying network models and empirical networks to highlight the role of different dynamical features on the spreading potential. The role of the observation time window is then analyzed in depth in order to provide indications on how this factor alters the estimated epidemic threshold.

II. DERIVATION OF THE EPIDEMIC THRESHOLD

We consider the susceptible-infected-susceptible (SIS) model [1] in discrete time, where individuals (i.e., the nodes of the network) can be in one of two mutually exclusive states—susceptible or infectious. At each time step, infectious individuals may transmit the infection to susceptible neighbors with probability λ along each contact, and they recover spontaneously with probability μ becoming susceptible once again. We consider the temporal network forming the substrate of the spreading process to be a sequence of undirected and unweighted static networks. The generalization of the following treatment to the directed and weighted case is outlined in Ref. [35].

In order to describe the spreading dynamics on such a substrate, we extend the Markov chain approach for static networks [8,9] to the case of temporal networks. The SIS propagation on a generic network with N nodes and adjacency matrix A is given by

$$p_i^{(t)} = 1 - \left[1 - (1-\mu)p_i^{(t-1)}\right]\prod_j\left[1 - \lambda A_{ji}p_j^{(t-1)}\right], \quad (1)$$

where $p_i^{(t)}$ is the probability for the node i to be in the infectious state at time t. The Markovian model of Eq. (1), widely adopted in different fields [12,36], is based on the mean-field assumption of the absence of dynamical

correlations among the states of neighboring nodes [37]. For both directed and undirected networks [38,39], the study of the asymptotic state yields the derivation of the epidemic threshold $(\lambda/\mu) = 1/\rho(A^\dagger)$, where $\rho(A^\dagger)$ is the spectral radius of the transposed adjacency matrix $A^\dagger$ [8,9]. This is known to be a lower bound estimate of the real epidemic threshold, approaching the real value with surprisingly high accuracy given the simplicity of the expression and its derivation [37,40].

We extend this paradigm to a temporal network by letting the adjacency matrix in Eq. (1) depend on time:

$$p_i^{(t)} = 1 - \left[1 - (1-\mu)p_i^{(t-1)}\right]\prod_j\left[1 - \lambda A_{ji}^{(t-1)}p_j^{(t-1)}\right]. \quad (2)$$

Here, $A^{(t)}$ is the adjacency matrix associated with the tth snapshot of the evolving network. In order to ensure the asymptotic solution of the SIS process in a generic temporal network, we assume periodic boundary conditions for the network dynamics. With T being the total number of network time snapshots, we impose $A^{(T+1)} \equiv A^{(1)}$. This does not imply any loss of generality given that T may be completely arbitrary. We notice that, as a consequence of the assumed periodic temporal dynamics of $A^{(t)}$, the asymptotic solution of Eq. (2) is, in principle, periodic, with period T.

We now define a more convenient representation of the coupled dynamics adopting the multilayer approach introduced in Ref. [33]. We map the temporal network to the tensor space $\mathbb{R}^N \otimes \mathbb{R}^T$, where each node is identified by the pair of indices (i, t), corresponding to the node label i and the time frame t, respectively. A multilayer representation of the temporal network can be introduced through the following rules:

   (i) Each node, at time t, is connected to its future self at $t+1$.
  (ii) If i is connected to j at time t, then we connect i at time t to j at time $t+1$, and j at time t to i at time $t+1$.

The second rule is termed "nondiagonal coupling" in the multilayer-network framework [34]. The first rule is consistent with the ordinal coupling in such a framework [34,41], but unlike in that representation, no links are found connecting nodes on the same layer since layers cease to correspond to the adjacency matrices of the temporal snapshots. The so-defined network is therefore multipartite since only pairs of nodes belonging to different layers are linked together (see Fig. 1 for a schematic illustration of this transformation). While formally falling into a specific subcase of the classification introduced in Ref. [42], the proposed mapping from the network temporal sequence to a multilayer object provides a novel representation of the temporal network that preserves the information relevant for the spreading process. The tensor representation of the obtained multilayer network is the following:
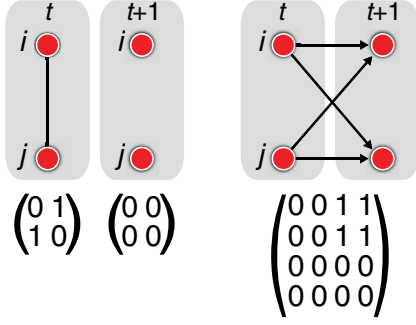
FIG. 1. Schematic example of the supra-adjacency matrix of the multilayer representation of the temporal network. For simplicity, we consider a network of two nodes i, j, and two time steps. The left panel represents the network as a sequence of static adjacency matrices. This is translated into a multilayer representation (right panel), where each node points to itself in the future and to the future image of its present neighbors.

$$A_{ij}^{tt'} = \delta^{t,t'+1}[\delta_{ij} + A_{ij}^{(t)}]. \tag{3}$$

Analogously to the definition of **A**, we can also write, in this representation, the tensor associated with the SIS dynamics of Eq. (2), coupling together contagion and network dynamics:

$$\mathbf{M}_{ij}^{tt'} = \delta^{t,t'+1}[(1-\mu)\delta_{ij} + \lambda A_{ij}^{(t)}]. \tag{4}$$

The multilayer representation and the definition of the tensor **M** introduce a simplified expression for Eq. (2). The tensor space can be represented in single-index notation through the isomorphism $\mathbb{R}^N \otimes \mathbb{R}^T \simeq \mathbb{R}^{NT}$. In other words, similarly to the definition of the supra-adjacency matrix in Refs. [33,43,44], we can mask the tensorial origin of the space through the map $(i, t) \to \alpha = Nt + i$, with α running in $\{1, ..., NT\}$, allowing us to write the network tensor **M** in matrix form:

$$M = \begin{pmatrix} 0 & 1-\mu+\lambda A^{(1)} & 0 & \cdots & 0 \\ 0 & 0 & 1-\mu+\lambda A^{(2)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-\mu+\lambda A^{(T-1)} \\ 1-\mu+\lambda A^{(T)} & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

M provides a network representation of the topological and temporal dimensions underlying the dynamics of Eq. (2), which are interrelated and flattened here. Its directed nature preserves the causality of the process, while its weights account for the SIS transition probabilities. The Markov process is now described by a trajectory in $\mathbb{R}^{NT}$ where the state vector $\hat{p}_\alpha(\tau)$ represents the probability of each node to be infected at each time step t included in the interval $[\tau T, (\tau+1)T]$. Consistently, Eq. (2) becomes

$$\hat{p}_\alpha(\tau) = 1 - \prod_\beta [1 - M_{\beta\alpha}\hat{p}_\beta(\tau-1)]. \tag{5}$$

Given that vector $\hat{p}$ encodes a one-period configuration, the T-periodic asymptotic state of the SIS process is now mapped into the steady state $\hat{p}_\alpha(\tau) = \hat{p}_\alpha(\tau-1)$. The latter can be recovered as a solution of the equilibrium equation:

$$\hat{p}_\alpha = 1 - \prod_\beta (1 - M_{\beta\alpha}\hat{p}_\beta), \tag{6}$$

which is formally the same as the stationary condition imposed in Eq. (1) for the static network case, and it is similar to the Markov chain approaches used to solve

contagion processes in multiplex and interconnected networks [43–45]. We can then follow Refs. [8,9] and linearize Eq. (6), recovering the necessary and sufficient condition for the asymptotically stable zero solution, $\rho(M^\dagger) < 1$ [46]. Considering that the uniform zero solution in the $\mathbb{R}^{NT}$ representation is mapped to a uniform zero solution in the original $\mathbb{R}^N$ representation, this yields the threshold condition

$$\rho(M^\dagger) = 1 \tag{7}$$

for the critical values of λ and μ above which the transmission becomes epidemic [8,9,43–45].

The spectral radius of M can be simplified with the following relation (see Appendix A):

$$\rho(M) = \rho(P)^{1/T}, \tag{8}$$

where $P = \prod_{t=1}^T (1-\mu+\lambda A^{(T-t)})$ represents a weighted version of the accessibility matrix [47], having connectivity weighted by λ and waiting time weighted by $1-\mu$. This last passage ensures a simplification of the numerical computation of the epidemic threshold, allowing an execution time scaling as $\sim TN^{5/2}$ (see Ref. [48] for an analysis of the numerical performance of our approach).

The quenched and annealed regimes can be recovered within this general framework as particular limiting solutions. In the first case, it is to be noted that the sequence of temporal snapshots naturally defines the minimum time scale of the process. In order to consider contagion dynamics that are much faster than the time-varying process of the network, we thus rely on the commonly adopted assumption regarding the temporal network as static, so $A^{(t)} \equiv A$. In this particular case, $P = (1 - \mu + \lambda A)^T$. Therefore, $\rho(M) = \rho(1 - \mu + \lambda A) = 1 - \mu + \lambda \rho(A)$. The requirement $\rho(M^\dagger) = 1$ thus recovers the expression known for the quenched case.

The study of the annealed regime is less trivial. In the assumption that λ and μ are very small, corresponding to a very slow disease dynamics with respect to the time scale of the network evolution, it is possible to replace P with its linear expansion in $\lambda/(1 - \mu)$, yielding

$$P_{\text{slow}} = (1 - \mu)^T \left[1 + \frac{\lambda}{1 - \mu} \mathcal{A} \right],$$

where $\mathcal{A} = \sum_t A^{(t)}$ is a weighted static representation of the network, formed by the sum of all the snapshots. Temporal correlations are lost, and edges count for the number of times they are active during the whole period T. Equation (7) for the epidemic threshold thus simplifies to $(\lambda/\mu)_{\text{c,slow}} = T/\rho(\mathcal{A})$ [49], and the aggregated matrix contains all the relevant information for spreading dynamics.

III. VALIDATION AND COMPARISON WITH STOCHASTIC SIMULATIONS

In the following, we validate the analytical method and compare its predictions with the behavior of a simulated SIS spreading process. For this purpose, we consider six networks, three of which are built from models for time-varying networks; the other three are obtained from empirical measures. The first network, ER, is formed by a sequence of random Erdős-Rényi graphs [50] with a given number of nodes and edges. It represents a simple and completely uncorrelated example of a temporal network. The second network, ACTIVITY, is a realization of the activity-driven model [28] where each node is assigned an activity potential, representing the probability of being active in a certain snapshot. Once activated, the node establishes a fixed number of connections that are renewed at each snapshot. We consider a heterogeneous activity distribution so that the obtained networks are characterized by a temporally uncorrelated sequence of snapshots with a heterogeneous aggregated degree distribution. BURSTY is built from the model introduced in Ref. [23] and accounts for a heterogeneous activation pattern describing a sequence of homogeneous networks where the intercontact time is power-law distributed. Size and period are chosen arbitrarily for all these networks since the choice of these parameters does not impact the method validation, as

discussed in more detail in the following section. As examples of real time-varying networks, we choose data sets describing different kinds of human contact: HT09 is the network of face-to-face proximity during a 2.5-day scientific conference [16]; SEX is a 1-year network of sexual contacts between prostitutes and their clients [18]; SCHOOL is a contact network describing one day in a high school [51]. Size, period, and topological properties are constrained by the measurements and are very diverse. Further information about the six networks can be found in Appendix B and in Table I.

To verify the validity of the proposed analytical expression, we numerically solve the Markov equation (2). For given λ and μ, we iterate the equation until the periodic state is reached and compute the average prevalence over a period $\langle i_{\text{MC}} \rangle = \sum_{i,t} p_i^{(t)}/(TN)$. Predictions are then also compared with the threshold behavior obtained from numerical simulations of the stochastic and microscopic SIS dynamics on the evolving networks. We use the quasistationary state method [52] (see Appendix C) to measure the average prevalence $\langle i_{\text{sim}} \rangle$ over the time series for different values of λ, after an initial transient time is discarded.

Figure 2 shows $\langle i_{\text{MC}} \rangle$ and $\langle i_{\text{sim}} \rangle$ as functions of λ for two different values of μ ($\mu = 0.2$ and $\mu = 0.5$) for all networks under study. The average prevalence displays the expected transition behavior. The solution of the Markov chain equation $\langle i_{\text{MC}} \rangle$ is equal to zero for small values of λ until the critical value of λ is reached, after which a rapid growth is observed signaling an epidemic affecting a finite fraction of the network. The transition is well predicted by the analytical expression of Eq. (7). The threshold behavior obtained from numerical simulations is also very similar to the mean-field prediction. The two curves of $\langle i_{\text{MC}} \rangle$ and $\langle i_{\text{sim}} \rangle$ are nearly superimposed, showing that the mean-field approximation in Eq. (2) is valid in all conditions of

TABLE I. Temporal networks considered for the validation. The first three networks are single realizations obtained from synthetic models for time-varying networks; the other three are empirical networks. The ER model is a sequence of random graphs with 500 nodes and 750 edges, so $\langle k \rangle = 3$. The ACTIVITY model is a sequence of snapshots built with parameter values: $\Delta t = 1$, $m = 2$, $\eta = 10$, $\gamma = 2.8$, $\epsilon = 3 \times 10^{-2}$, in the notation of Ref. [28]. The BURSTY network is built with a power-law distributed interactivation time, with exponent -2, and cutoff equal to the period of the network. For the real networks, the collection time is the total time considered in the data set.

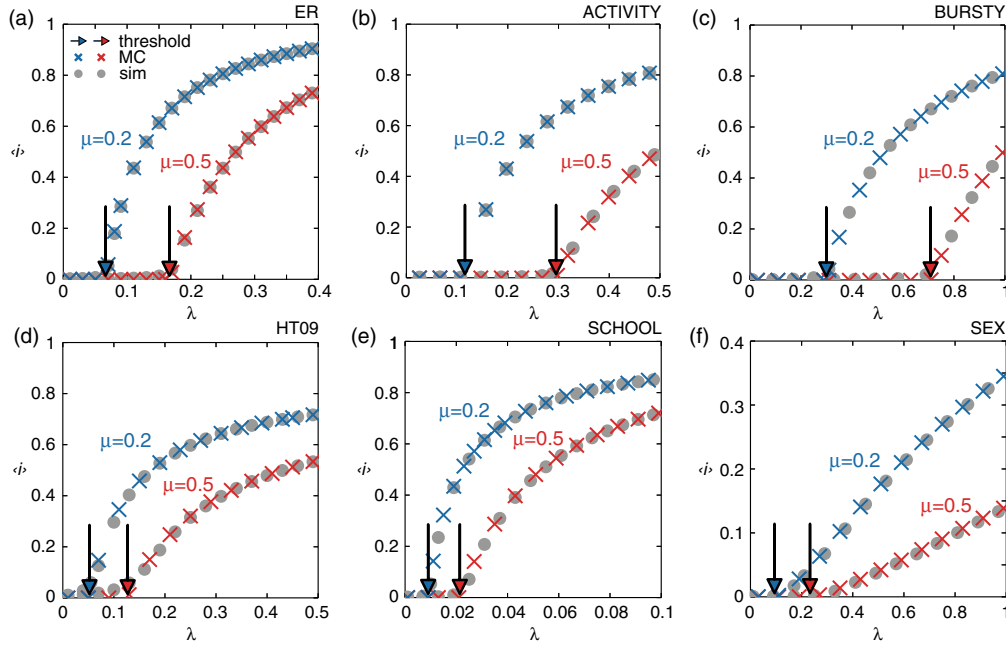| Network	Number of nodes	Period T	Aggregating window
ER [50]	500	13	$\cdots$
ACTIVITY [28]	1000	20	$\cdots$
BURSTY [23]	500	50	$\cdots$
HT09 [16]	113	30	1 hour
SCHOOL [51]	787	42	10 mins
SEX [18]	6866	13	28 days

FIG. 2.   Validation of the analytical method and comparison with microscopic numerical simulations. Top panel: Network models. Bottom panel: Empirical networks. Cross symbols represent $\langle i_{MC} \rangle$ as a function of λ for two different values of μ ($\mu = 0.2$ in blue and $\mu = 0.5$ in red), i.e., the average prevalence obtained from the numerical solution of the Markov chain, Eq. (2). Circles represent $\langle i_{sim} \rangle$, i.e., the average prevalence obtained from stochastic microscopic numerical simulations of the SIS process, for the same values of μ. The arrows indicate the analytical predictions of the threshold from Eq. (8) (one single realization of the network models is considered in each panel).

network size and average connectivity here considered. The presence of correlations shows its effects in proximity to the transition, which is smoother for $\langle i_{sim} \rangle$ with respect to $\langle i_{MC} \rangle$. This is particularly evident for the network HT09 and is a consequence of its small size ($N = 113$). In Ref. [53], we report the analysis of the dynamical correlations.

The good agreement between the computed epidemic threshold, the solution of the Markov chain equation, and the numerical simulation results is thus maintained under a range of different temporal network properties (presence or absence of temporal correlations, heterogeneous vs homogeneous distributions characterizing temporal and structural observables, and the possible presence of community structure as in the case of school) and sizes (from approximately 10^2 nodes to 10^4). It is important to mention that periodic boundary conditions in the case of real networks may, in principle, induce nonexisting phenomena (such as, for example, temporal paths [54]) that could alter the threshold estimation by influencing the spreading process. We analyze the effect of our technical assumption of adopting periodic constraints in the following section, also in relation to data availability and collection.

IV. OPTIMAL DATA COLLECTION TIME

Available data sets characterizing empirical networks only account for a portion of the real contact process, and

the extent of the recording time window may affect the prediction of the epidemic threshold. One may expect that, when the data-collection period is long enough, the data would represent an approximately complete reconstruction of the temporal network properties, thus leading to an accurate estimate of the epidemic threshold. Given the resources needed for the setup of data-collection deployments, here we explore the role of the period T aimed at identifying a minimum length of observation of the contact process that is optimal in providing a reliable characterization of the spreading potential.

We thus compute the epidemic threshold from Eq. (8) for increasingly larger values of the period T up to the entire data-collection time window, for the three empirical networks under study. Figure 3 shows a saturation behavior for λ_c, indicating that the data-collection period is long enough to characterize the epidemic dynamics. Such behavior and its associated relaxation time strongly depend on the network's typical time scale and on the temporal variability of its structure. In more detail, a simple structural measure— the variation of the average degree along the period—is shown to strongly impact the predicted λ_c (Fig. 3 and Ref. [55]). This is particularly evident in the SCHOOL network, where the daily activity of students determines considerable variations in the average degree and induces marked oscillations in the resulting threshold. In addition to the empirical networks, we also consider the BURSTY network model as it includes nontrivial temporal correlations.
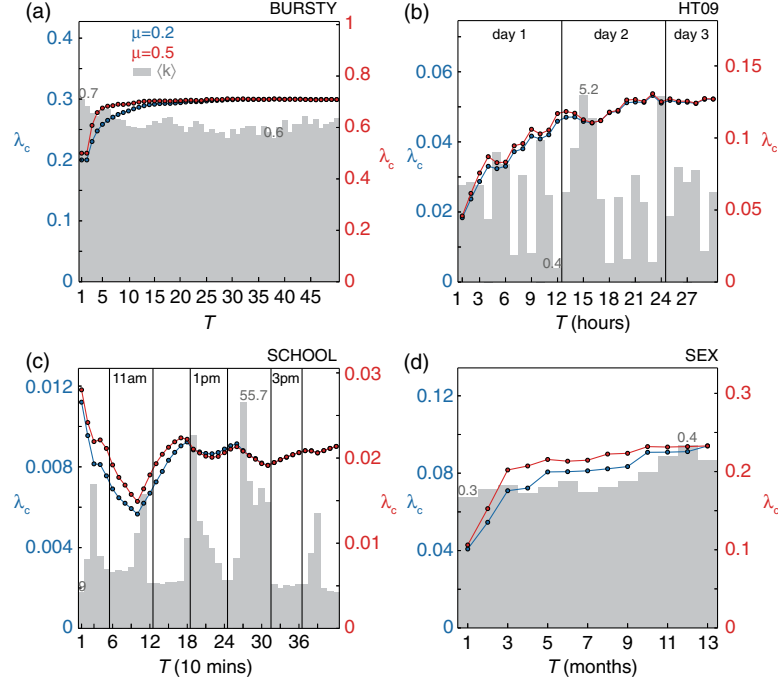
FIG. 3.   Epidemic threshold estimated from different period lengths. $\lambda_c(T)$ is the epidemic threshold computed by considering only the first T snapshots of the network. For each panel, the blue (red) curve corresponds to $\mu = 0.2$ ($\mu = 0.5$), and its scale is reported on the left (right) side of the plot. The gray bar chart shows the mean network degree associated with the snapshot at time T. Bar charts are in linear scale, and min and max values are placed on the corresponding bars. For the empirical networks, the measure of the real time is also reported.

In this case, the critical transmission probability λ_c rapidly saturates to a constant value (Fig. 3), and an even more rapid saturation is observed for the other two network models (see Ref. [56]). The average degree is indeed relatively stable, so small temporal windows are enough to fully characterize λ_c.

Different values of the recovery probability μ lead, in general, to similar behaviors of λ_c towards saturation, differing essentially by a scaling factor. The effect of μ on saturation time is instead visible for the BURSTY network. In this case, when the period length is smaller than the average duration of the infection, the truncation in the intercontact time distribution clearly alters the estimation of the threshold.

These results indicate that it is possible to identify a minimum length of the observation window of a real system for contact data collection, highlighting the presence of well-defined properties and patterns characterizing the system that can be captured in a finite time.

V. CONCLUSION

Being able to provide a reliable and accurate estimation of the epidemic threshold for a spreading process taking place on a given networked system is of the utmost importance, as it allows predictions of the likelihood of a wide-spreading event and identification of containment measures (crucial for infectious disease epidemics) or strategies for enhancing the propagation (desired in the

case of information diffusion). While analytical approaches have so far targeted only specific contexts, our framework allows the analytical computation of the epidemic threshold on an arbitrary temporal network, requiring no assumption about the network topology or time variation. The proposed approach is based on the spectral decomposition of the flattened matrix representation of the topological and temporal structure of the network, extending the Markov chain model introduced for the static network case to its temporal counterpart. The predicted epidemic threshold, validated against the numerical solution of the model, also reproduces the behavior observed in stochastic microscopic numerical simulations of the spreading process with high accuracy.

The technical requirement of periodic conditions does not limit the general applicability of our approach, as the method is valid for an arbitrary period length. Moreover, this feature allows us to inform data-collection endeavors on the time period of observation of the system required to fully characterize its spreading properties. Our focus on the discrete time formulation of the process is prompted by the study of several empirical networks for which time is naturally discrete, the time step being dictated by the resolution of the data-collection procedure. Extensions to the continuous-time case would be needed when the continuous-time description is more appropriate, as for example with some modeling approaches.

Our framework thus introduces a multilayer formulation of spreading phenomena on time-varying networks that opens the path to new theoretical understandings of the complex interplay between the two temporal processes, disentangling the role of the network's dynamical features, such as activation rate, temporal correlations, and temporal resolution.

APPENDIX A: PROOF OF EQ. (8)

Computing the eigenvalues of $M^\dagger$ means solving the equation $\det(x - M^\dagger) = 0$, where the determinant is computed on the $\mathbb{R}^{NT}$ space ($\det_{NT}$). Given that $x - M^\dagger$ is composed of T^2 blocks of size $N \times N$, we can use the findings in Ref. [57] to reduce the dimensionality of the problem, i.e., $\det_{NT} \rightarrow \det_N$. Moreover, given that several blocks of $x - M^\dagger$ are zero, the general result in Ref. [57] simplifies to $\det_{NT}(x - M^\dagger) = (-1)^{NT} \det_N(x^T - P)$. Equation (8) immediately follows.

APPENDIX B: NETWORKS CONSIDERED FOR VALIDATION

In this section, we provide the details of the six networks used for the validation of the threshold expression.

ER: The network is formed by a sequence of random Erdős-Rényi graphs [50] with 500 nodes and 750 edges, so $\langle k \rangle = 3$.

ACTIVITY: In the activity-driven model by Perra *et al.* [28], nodes that are active in a certain snapshot establish a fixed number of connections (in our case, 2) with other nodes picked at random (both active and inactive). All links are renewed after every snapshot. The activity potential is assigned by sampling numbers $x \in [\epsilon, 1]$ from a power-law distribution (in our case, with exponent 2.8 and $\epsilon = 3 \times 10^{-2}$) and then converting them to activity potentials $a = 1 - e^{-\eta x}$. η is a free parameter used to tune the average degree; here, $\eta = 10$.

BURSTY: The network is built from the model introduced by Rocha *et al.* [23], where a node becomes active at time t with a probability that depends on the time it was last active, t'. If active, it then forms a link with another active node. All links are removed before proceeding to the following snapshots. To enforce a BURSTY interevent time distribution, the probability of becoming active is sampled with the distribution $(t - t')^{-\alpha_1} e^{-\alpha_2(t-t')}$, with $\alpha_1 = 2$ and $\alpha_2 = 5 \times 10^{-4}$.

HT09: The data set was collected by the Sociopatterns group [16], and it records the interactions among participants at a scientific conference. Links represent face-to-face proximity recorded by wearable radio-frequency identification (RFID) tags. Time resolution of the signal is 20 sec. Each tag emits wave packets that may be recorded by other tags, thus signaling proximity. Tags were embedded in conference badges; their signal intensity was set to be detectable at a maximum distance of 1 m and completely shielded by the human body. With these settings in effect, only close proximity in a face-to-face position resulted in interaction [58].

SEX: This is a network of sexual contacts between female prostitutes and their male clients as inferred from posts on a Brazilian online escort forum where customers could rate their experience with a certain sex worker. The date of the post was taken as a proxy for the time of the sexual intercourse. The (anonymized) data set can be found in Ref. [18].

SCHOOL: This network represents the face-to-face proximity interactions during a day in a high school [51]. Students and staff were given wearable RFID sensors, and proximity was recorded in a similar fashion as for HT09 [59].

APPENDIX C: ESTIMATION OF THE EPIDEMIC THRESHOLD FROM NUMERICAL SIMULATIONS

The computation of $\langle i_{\text{sim}} \rangle$ in proximity to the transition is difficult because surviving configurations are rare and a very large number of realizations of the process is needed to collect substantial statistics. We use the quasistationary state (QS) method [52,60] to overcome this difficulty and increase our computational efficiency. The QS method is based on the idea of constraining the system in an active state. Every time the absorbing state (i.e., no infected) is reached by the system, it is substituted with an active configuration that is randomly taken from the history of the simulation. In particular, 50 active configurations for each network snapshot are kept in memory. Whenever an active configuration is reached, it replaces one of the 50 with probability 0.2. When the absorbing state is reached, an active configuration is chosen among these 50 of that particular snapshot. For each simulation, after a relaxation time of 3×10^3 time steps, statistics are collected during 10^5 time steps. The method produces a time series that is long enough to accurately compute the observables $\langle i_{\text{sim}} \rangle$.

[1] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, 2008).

[2] V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani, *Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions*, PLoS Med. **4**, e13 (2007).

[3] A. J. Tatem and D. L. Smith, *International Population Movements and Regional Plasmodium Falciparum Malaria Elimination Strategies*, Proc. Natl. Acad. Sci. U.S.A. **107**, 12222 (2010).

[4] R. Pastor-Satorras and A. Vespignani, *Epidemic Spreading in Scale-Free Networks*, Phys. Rev. Lett. **86**, 3200 (2001).

[5] W. Goffman and V. A. Newill, *Generalization of Epidemic Theory: An Application to the Transmission of Ideas*, Nature (London) **204**, 225 (1964).

[6] D. J. Daley and D. G. Kendall, *Epidemics and Rumours*, Nature (London) **204**, 1118 (1964).

[7] S. Bikhchandani, D. Hirshleifer, and I. Welch, *A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades*, J. Polit. Econ. **100**, 992 (1992).

[8] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, *Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint*, in *Proceedings of the 22nd International Symposium on Reliable Distributed Systems* (IEEE, Los Alamitos, CA, 2003), pp. 25–34.

[9] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, *Discrete Time Markov Chain Approach to Contact-Based Disease Spreading in Complex Networks*, Europhys. Lett. **89**, 38009 (2010).

[10] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, *Resilience of the Internet to Random Breakdowns*, Phys. Rev. Lett. **85**, 4626 (2000).

[11] M. E. J. Newman, *Spread of Epidemic Disease on Networks*, Phys. Rev. E **66**, 016128 (2002).

[12] M. Boguñá, C. Castellano, and R. Pastor-Satorras, *Nature of the Epidemic Threshold for the Susceptible-Infected-Susceptible Dynamics in Networks*, Phys. Rev. Lett. **111**, 068701 (2013).

[13] C. Castellano and R. Pastor-Satorras, *Thresholds for Epidemic Spreading in Networks*, Phys. Rev. Lett. **105**, 218701 (2010).

[14] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes, *Localization and Spreading of Diseases in Complex Networks*, Phys. Rev. Lett. **109**, 128702 (2012).

[15] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).

[16] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, *What's in a Crowd? Analysis of Face-to-Face Behavioral Networks*, J. Theor. Biol. **271**, 166 (2011).

[17] G. Miritello, E. Moro, and R. Lara, *Dynamical Strength of Social Ties in Information Spreading*, Phys. Rev. E **83**, 045102 (2011).

[18] L. E. C. Rocha, F. Liljeros, and P. Holme, *Information Dynamics Shape the Sexual Networks of Internet-Mediated Prostitution*, Proc. Natl. Acad. Sci. U.S.A. **107**, 5706 (2010).

[19] P. Bajardi, A. Barrat, F. Natale, L. Savini, and V. Colizza, *Dynamical Patterns of Cattle Trade Movements*, PLoS One **6**, e19869 (2011).

[20] C. T. Butts, *Revisiting the Foundations of Network Analysis*, Science **325**, 414 (2009).

[21] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, *Small but Slow World: How Network Topology and Burstiness Slow Down Spreading*, Phys. Rev. E **83**, 025102 (2011).

[22] P. Holme and J. Saramäki, *Temporal Networks*, Phys. Rep. **519**, 97 (2012).

[23] L. E. C. Rocha and V. D. Blondel, *Bursts of Vertex Activation and Epidemics in Evolving Networks*, PLoS Comput. Biol. **9**, e1002974 (2013).

[24] L. Ferreri, P. Bajardi, M. Giacobini, S. Perazzo, and E. Venturino, *Interplay of Network Dynamics and Heterogeneity of Ties on Spreading Dynamics*, Phys. Rev. E **90**, 012812 (2014).

[25] J. Luis Iribarren and E. Moro, *Impact of Human Activity Patterns on the Dynamics of Information Diffusion*, Phys. Rev. Lett. **103**, 038702 (2009).

[26] A. Vazquez, B. Rácz, A. Lukács, and A.-L. Barabási, *Impact of Non-Poissonian Activity Patterns on Spreading Processes*, Phys. Rev. Lett. **98**, 158702 (2007).

[27] K. T. D. Eames and M. J. Keeling, *Monogamous Networks and the Spread of Sexually Transmitted Diseases*, Math. Biosci. **189**, 115 (2004).

[28] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, *Activity Driven Modeling of Time Varying Networks*, Sci. Rep. **2**, 469 (2012).

[29] E. Volz and L. Ancel Meyers, *Epidemic Thresholds in Dynamic Contact Networks*, J. R. Soc. Interface **6**, 233 (2009).

[30] T. Gross, C. J. Dommar D'Lima, and B. Blasius, *Epidemic Dynamics on an Adaptive Network*, Phys. Rev. Lett. **96**, 208701 (2006).

[31] M. Taylor, T. J. Taylor, and I. Z. Kiss, *Epidemic Threshold and Control in a Dynamic Network*, Phys. Rev. E **85**, 016103 (2012).

[32] S. Liu, N. Perra, M. Karsai, and A. Vespignani, *Controlling Contagion Processes in Activity Driven Networks*, Phys. Rev. Lett. **112**, 118702 (2014).

[33] M. De Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, *Mathematical Formulation of Multilayer Networks*, Phys. Rev. X **3**, 041022 (2013).

[34] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, *Multilayer Networks*, J. Comp. Networks **2(3)**, 203 (2014).

[35] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.5.021005 for an extension to the weighted and directed cases.

[36] P. Van Mieghem, *The N-Intertwined SIS Epidemic Network Model*, Computing **93**, 147 (2011).

[37] J. P. Gleeson, S. Melnik, J. A. Ward, M. A. Porter, and P. J. Mucha, *Accuracy of Mean-Field Theory for Dynamics on Real-World Networks*, Phys. Rev. E **85**, 026106 (2012).

[38] P. Van Mieghem and R. van de Bovenkamp, *Non-Markovian Infection Spread Dramatically Alters the*

Susceptible-Infected-Susceptible Epidemic Threshold in Networks, Phys. Rev. Lett. **110**, 108701 (2013).

[39] C. Li, H. Wang, and P. Van Mieghem, *Epidemic Threshold in Directed Networks*, Phys. Rev. E **88**, 062802 (2013).

[40] E. Cator and P. Van Mieghem, *Second-Order Mean-Field Susceptible-Infected-Susceptible Epidemic Threshold*, Phys. Rev. E **85**, 056111 (2012).

[41] P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, *Community Structure in Time-Dependent, Multiscale, and Multiplex Networks*, Science **329**, 277 (2010).

[42] K. Wehmuth, A. Ziviani, and E. Fleury, *A Unifying Model for Representing Time-Varying Graphs*, arXiv: 1402.3488.

[43] E. Cozzo, R. A. Baños, S. Meloni, and Y. Moreno, *Contact-Based Social Contagion in Multiplex Networks*, Phys. Rev. E **88**, 050801 (2013).

[44] H. Wang, Q. Li, G. D'Agostino, S. Havlin, H. Eugene Stanley, and P. Van Mieghem, *Effect of the Interconnected Network Structure on the Epidemic Threshold*, Phys. Rev. E **88**, 022801 (2013).

[45] C. Granell, S. Gómez, and A. Arenas, *Dynamical Interplay between Awareness and Epidemic Spreading in Multiplex Networks*, Phys. Rev. Lett. **111**, 128701 (2013).

[46] S. Elaydi, *An Introduction to Difference Equations*, 3rd ed. (Springer, New York, NY, 2005).

[47] H. H. K. Lentz, T. Selhorst, and I. M. Sokolov, *Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks*, Phys. Rev. Lett. **110**, 118701 (2013).

[48] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.5.021005 for performance of the algorithm used to compute the threshold.

[49] N. Masuda, K. Klemm, and V. M. Eguiluz, *Temporal Networks: Slowing Down Diffusion by Long Lasting Interactions*, Phys. Rev. Lett. **111**, 188701 (2013).

[50] P. Erdős and A. Rényi, *On the Evolution of Random Graphs*, Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17 (1960).

[51] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, *A High-Resolution Human Contact Network for Infectious Disease Transmission*, Proc. Natl. Acad. Sci. U.S.A. **107**, 22020 (2010).

[52] S. C. Ferreira, C. Castellano, and R. Pastor-Satorras, *Epidemic Thresholds of the Susceptible-Infected-Susceptible Model on Networks: A Comparison of Numerical and Theoretical Results.*, Phys. Rev. E **86**, 041125 (2012).

[53] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.5.021005 for an analysis of dynamical correlations.

[54] R. K. Pan and J. Saramäki, *Path Lengths, Correlations, and Centrality in Temporal Networks*, Phys. Rev. E **84**, 016105 (2011).

[55] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.5.021005 for further plots showing the impact of average degree variations on threshold fluctuations.

[56] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevX.5.021005 for the saturation of the threshold in the other network models.

[57] P. D. Powell, *Calculating Determinants of Block Matrices*, arXiv:1112.4379.

[58] Data sets and additional information are available on www.sociopatterns.org.

[59] Data can be found on www.salathegroup.com.

[60] K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Physics, An Introduction* (Springer, Berlin, 2010).

Analytical computation of the epidemic threshold on temporal networks

Supplemental Material

Eugenio Valdano[1,2], Luca Ferreri[3] Chiara Poletto[1,2],
and Vittoria Colizza[1,2,4]

[1]INSERM, UMR-S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France.
[2]Sorbonne Universités, UPMC Univ Paris 06, UMR-S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France.
[3]Dipartimento di Scienze Veterinarie, Università degli Studi di Torino, Grugliasco (TO), Italy.
[4]ISI Foundation, Torino, Italy.

1 Optimal data collection time

1.1 ER and ACTIVITY

Fig. S1 completes the picture given in Fig. 3 of main paper, by showing how period length T affects the epidemic threshold for the models ER and ACTIVITY. In both cases the threshold converges very quickly to a constant value, so that for $T > 5$ there are no significant oscillations. This is associated with low variations in the average degree among the different snapshots (variations being exactly equal to zero in the ER network by constructions). A small number of snapshots needs to be generated in order to correctly compute the epidemic threshold.

1.2 More on the correlation between threshold and degree fluctuations

In main paper we show how the oscillations of λ_c are associated with the instantaneous fluctuations of the average degree of the network, as the period T varies. As T increases, λ_c fluctuations are damped, because they are cumulatively calculated on longer periods. Adding a snapshot to a short period T has indeed a greater relative influence on the threshold than adding it to an already long period. Similarly, damped oscillations are observed when we monitor the cumulative average degree of the network, calculated on all snapshots up to period T (Figure S2).
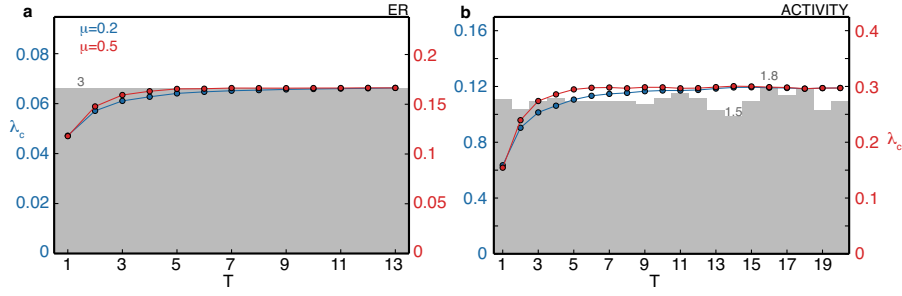
Figure S1: **Epidemic threshold estimated from different period lengths.** $\lambda_c(T)$ is the epidemic threshold computed by considering only the first T snapshots of the network. For each panel, blue (red) curve corresponds to $\mu = 0.2$ ($\mu = 0.5$). The scale of the former is on the left side of the plot, while the scale of the latter is on the right. The gray bar chart shows the mean network degree associated to the snapshot at time T. Bar charts are in linear scale and min and max values are placed on the corresponding bars.
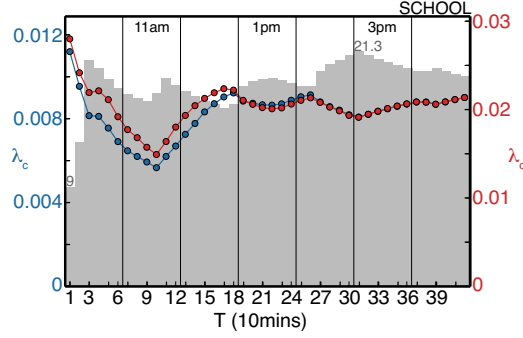
Figure S2: **Epidemic threshold estimated from different period lengths.** $\lambda_c(T)$ is the epidemic threshold computed by considering only the first T snapshots of the network. For each panel, blue (red) curve corresponds to $\mu = 0.2$ ($\mu = 0.5$). The scale of the former is on the left side of the plot, while the scale of the latter is on the right. The gray bar chart shows the average of the mean network degree of snapshots up to time T. Bar charts are in linear scale and min and max values are placed on the corresponding bars. The measure of the real time is also reported.

2 Higher order correlations

The quenched mean field approach disregards spatial correlations among the probabilities of nodes being infected. Let $X_i(t) = 0, 1$ be the infectious status of node i at time t. The quenched mean field then assumes that $\langle X_i X_j \rangle = \langle X_i \rangle \langle X_j \rangle$. In order to assess the impact of such approximation, at least in the case of two point correlations, we compute in the simulations the following quantity for each pair of nodes:

$$\sigma_{ij} = \frac{\langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle}{\sqrt{\langle X_i^2 \rangle - \langle X_i \rangle^2} \sqrt{\langle X_j^2 \rangle - \langle X_j \rangle^2}}, \tag{1}$$

where the moments are computed on a time interval well after the initial transient: $\langle Y \rangle = \left[\sum_{t=t_0}^{t_1} Y(t) \right] / (t_1 - t_0)$. $\sigma_{ij} \approx 1$ indicates that X_i, X_j are highly correlated, while low values of that quantity indicate no correlations. Figure S3 shows the distribution for σ_{ij} on all possible pairs of nodes, for ER and HT09. For ER almost no correlations are visible among nodes, as expected, except for small fluctuations around $\sigma_{ij} = 0$. For HT09, on the other hand, we see that σ_{ij} is peaked around a small but nonzero value, indicating the presence of two point correlations, albeit weak. This is in accordance with the finding that the threshold computation is more accurate for ER than for HT09 (see Figure 2a,d of main paper).
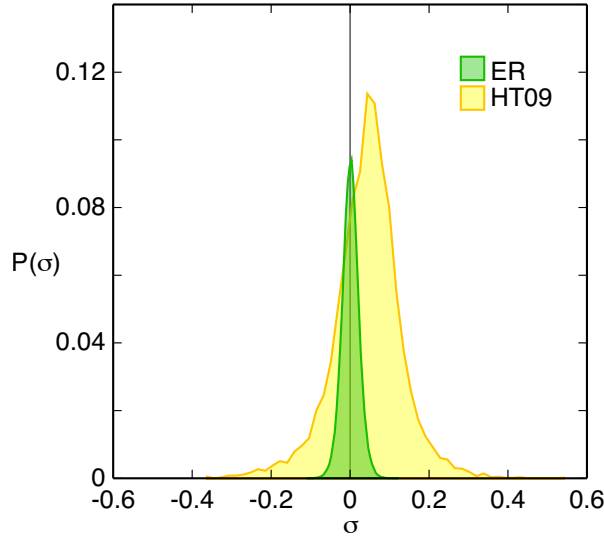
Figure S3: **Two-point correlations for ER and HT09.** Distribution of $\sigma = \sigma_{ij}$ for the model ER and the real network HT09. Distributions are computed using $\lambda = 0.3$, $\mu = 0.5$ for both networks. Moments are computed over $2 \cdot 10^5$ time steps, after a relaxation time of $3 \cdot 10^3$.

3 Weighted and directed networks

For sake of simplicity, in the main paper we deal with undirected unweighted temporal networks, i.e. $A^{(t)\dagger} = A^{(t)}$ and $A_{ij}^{(t)} \in \{0, 1\}$. Here we briefly show that our methodology can be extended to the more general case, if needed.

The proposed approach does not require $A^{(t)}$ to be symmetric (moreover $M_{\alpha\beta}$ is not symmetric even when the $A^{(t)}$ matrices are). Therefore a generalisation to the directed case is straightforward, and it is simply based on the replacement of $\mathcal{A}$ with $\mathcal{A}^\dagger$ in the computation of the threshold in the annealed approximation.

The weighted case is tractable too, provided that the probability of transmitting the infection is defined in terms of the weight, according to the given context under study (the weight could for example represent the movements of hosts from one node to another).

It is customary to compute the probability of transmission along a link ij as $1 - (1 - \lambda)^{A_{ij}^{(t)}}$ (see, for instance, [1]). When the weights are integer numbers, this means

3

considering the binomially distributed probability of at least one infection given $A_{ij}^{(t)}$ trials. By plugging this contribution into $\mathbf{M}_{ij}^{tt'}$, equation (4) of the main paper thus becomes

$$\mathbf{M}_{ij}^{tt'} = \delta^{t,t'+1} \left[(1-\mu)\, \delta_{ij} + 1 - (1-\lambda)^{A_{ij}^{(t)}} \right].\qquad(2)$$

The computation of the threshold can then be carried out in the same way as described in main paper.

4 Computational performance

In this section we discuss the performance and scalability of the numerical algorithm we use to compute the spectral radius of matrix P. The algorithm is implemented in `Python 2.7` and uses `numpy` and `scipy` libraries for sparse matrix representation and multiplication, `numpy.dot` and `scipy.sparse.csr`. The spectral radius of P is computed through a modified version of the well known power iteration method [2]. We tested the scaling of the execution time by varying the number of nodes N and the period T of an uncorrelated sequence of Erdős-Rény with $\langle k \rangle = 2$. The execution time grows linearly with the period and as the $5/2$-th power of the number of nodes, $t \sim T N^{5/2}$. As a point of reference, the average execution time for $N = 10^3$ and $T = 10^2$ is 15 seconds on an Intel Core i5 3.2G Hz with RAM DDR3 8 GB and 1.6 GHz.

SM References

[1] Paolo Bajardi, Alain Barrat, Lara Savini & Vittoria Colizza, Optimizing surveillance for livestock disease spreading through animal movements. *J. R. Soc. Interface*. doi:10.1098/rsif.2012.0289 (2012).

[2] Herman Müntz, Solution directe de l'équation séculaire et de quelques problèmes analogues transcendants, *C. R. Acad. Sci. Paris* 156,43-46 (1913).

END

3.3. Article: Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution

In many real scenarios, like cattle trade networks (Chapter 4), contacts are modeled through weighted, directed links, to account for non-homogeneous one-directional force of infection. In addition, the SIS model is not the adequate choice for diseases that exhibit immunity after recovery. In the following article we extend the calculation of the threshold to weighted, directed networks, and to the SIR and SIRS compartmental models (for their description see Figure 1.1). We apply the approach to two temporal network models and one empirical dataset. We find that immunity and loss of immunity, features of these new compartmental models, do not affect the estimation of the epidemic threshold through the infection propagator approach. Furthermore, we show that aggregating the temporal network causes the estimated epidemic threshold to rapidly lose accuracy. We test both topological and weighted aggregation schemes (see Chapter 2), and find that weight-topology correlations are found to be the critical factor to be preserved to improve accuracy in the prediction.

BEGINNING

Valdano E, Poletto C, Colizza V

Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution

under review (2015)

Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution

Eugenio Valdano[1], Chiara Poletto[1],
and Vittoria Colizza[1,2]

[1]Sorbonne Universités, UPMC Univ Paris 06, INSERM, Institut Pierre Louis d'Épidémiologie et de Santé
Publique (IPLESP UMRS 1136), F75012, Paris, France
[2]ISI Foundation, Torino, Italy.

Abstract

The epidemic threshold of a spreading process indicates the condition for the occurrence of the wide spreading regime, thus representing a predictor of the network vulnerability to the epidemic. Such threshold depends on the natural history of the disease and on the pattern of contacts of the network with its time variation. Based on the theoretical framework introduced in (Valdano et al. PRX 2015) for a susceptible-infectious-susceptible model, we formulate here an infection propagator approach to compute the epidemic threshold accounting for more realistic effects regarding a varying force of infection per contact, the presence of immunity, and a limited time resolution of the temporal network. We apply the approach to two temporal network models and an empirical dataset of school contacts. We find that immunity and loss of immunity do not affect the estimation of the epidemic threshold through the infection propagator approach. Comparisons with numerical results show the good agreement of the analytical predictions. Aggregating the temporal network rapidly deteriorates the predictions, except for slow diseases once the heterogeneity of the links is preserved. Weight-topology correlations are found to be the critical factor to be preserved to improve accuracy in the prediction.

1 Introduction

The concept of epidemic threshold is fundamental in infectious disease modeling [1, 2]. When a pathogen is seeded in a population, a critical transmissibility exists below which the spread rapidly ceases. Such a threshold is a combined property of the disease natural history and of the network of interactions along which transmission can occur. In the physics literature such interplay has been typically studied for the family of susceptible-infected-susceptible and susceptible-infected-recoeverd models on networks [3–11]. Several analytical approaches based, for instance, on the heterogenous mean field approximation [3], on percolation theory [6, 7] and on Markov processes [4, 5] have been developed to study the transition from early extinction to epidemic.

Extensive work has been done under the assumption of spreading time scales either much slower or much faster than the one characteristic of the underlying network – the two regimes called annealed and quenched, respectively [3,5]. In recent years, the massive amount of empirical information on networks has showed that such assumption does not hold in many cases [12–19] and that the network dynamics presents features (e.g. memory, bursty activation, heterogeneities in node activity) affecting the resulting spreading processes [13, 14, 18, 20–28].

The majority of studies addressing so far the impact of network dynamics on the epidemic spread through the analytical calculation of the epidemic threshold are all based on synthetic models of the network evolution, valid under context-specific assumptions [21, 24–29]. To fill this gap, we have introduced in [30] a method to compute the epidemic threshold for a susceptible-infectious-susceptible (SIS) process on a generic discrete-time temporal network, assuming the knowledge of its sequence of adjacency matrices. The approach is rooted in a multi-layer representation [31, 32] of the temporal network that preserves the network causality. It employs a tensor formulation that integrates both spreading and network dynamics and allows for the analytical solution of the linearized Markov chain description of the spreading process. Such framework extends in this way the quenched approach to the time-varying case, through a multilayer transformation.

The lack of assumption on the network substrate makes such a tool a candidate for assessing the vulnerability to epidemic invasion of real systems for which time-varying contact data relevant for epidemic transmissions are collected [13–19, 33–36]. At the same time, it allows a systematic exploration of the structural and temporal factors characterizing the time-evolving network that are responsible for sustained spreading. To allow the use of this framework to a variety of different settings and epidemic conditions, we assess here the applicability of the approach in describing realistic diseases and its robustness with respect to network properties induced by data collection procedures and availability. By considering empirical and synthetic model contact data, we discuss how a varying force of infection along a given link and its direction impact the computation of the threshold for a SIS dynamics. Then, we formulate the approach for more realistic disease natural histories, considering susceptible-infectious-recovered (SIR) and susceptible-infectious-recovered-susceptible (SIRS) compartmental models. This allows us to account for an additional ingredient – immunity following infection, either permanent or temporary – representing an important key feature of many diseases. Finally, we address the problem of limited temporal resolution in the knowledge or availability of the network dynamics, for which contacts occurring within a given time interval are

aggregated [37]. By focusing on an empirical network of time-varying contacts among individuals at school, we quantify the accuracy and reliability of the estimation of the epidemic threshold testing increasing aggregations, to provide quantitative and qualitative information on the specific temporal structures responsible for observed biases.

2 Infection propagator approach for a weighted directed temporal network

We consider a SIS model [1, 2] where hosts, represented by nodes in the network, can be either in the susceptible or the infectious state. We assume the process to unfold in discrete time on a weighted directed temporal network, comprising a finite number T of snapshots, each one with a weighted adjacency matrix $\mathbf{W}_t$. The entry $W_{t,ij}$ encodes the weight of the directed link from i to j at time step t. At each time step, infectious nodes spontaneously recover with probability μ, returning to the susceptible state. While infectious, nodes can transmit the infection to susceptible neighbors with a probability that depends both on the weight of the link and on the intrinsic transmissibility of the pathogen λ, representing the probability of transmission when link weight is equal to 1. We model this by introducing a transmission matrix $\mathbf{\Lambda}_t$, function of both λ and $\mathbf{W}_t$, that encodes transmission probabilities. When the network is unweighted ($W_{t,ij} = A_{t,ij} = 0, 1$), the entries of the matrix $\mathbf{\Lambda}_t$ are simply given by $\Lambda_{t,ij} = \lambda A_{t,ij}$. If the network is weighted, several choices are possible to model transmission along the weighted link. Here we consider a binomial process for the infection, so that:

$$\Lambda_{t,ij} = 1 - (1 - \lambda)^{W_{t,ij}}, \tag{1}$$

as it is typically assumed, for example, in the spread of livestock infections between premises where the weight represent the number of animals moved between farms [38].

We start from the microscopic Markov chain approach, or *quenched mean field approach*, developed for static networks [4, 5]. According to this, the equations describing the SIS propagation on a generic static network with N nodes and adjacency matrix $\mathbf{W}$ are

$$p_{t,i} = 1 - [1 - (1 - \mu) p_{t-1,i}] \prod_j (1 - \Lambda_{ji} p_{t-1,j}), \tag{2}$$

where $p_{t,i}$ is the probability that node i is in the infectious state at time t, and $\mathbf{\Lambda}$ is the static transmission matrix. We remark that Eq. (2) relies on the assumption that no dynamical correlations exist among infection probabilities of neighbouring nodes [39].

The microscopic Markov chain model of Eq. (2) is widely adopted in different fields [8, 40]. For both directed and undirected networks [41, 42] the study of its asymptotic state yields the derivation of the epidemic threshold in terms of the spectral radius of the transmission matrix $\rho[\mathbf{\Lambda}]$, namely the modulus of the largest eigenvalue of $\mathbf{\Lambda}$. The threshold is the value of λ for which the following holds: $\rho[\mathbf{\Lambda}] = \mu$. In the unweighted case, this is equivalent to the well-known relation $(\lambda/\mu) = 1/\rho[\mathbf{A}]$, where $\rho[\mathbf{A}]$ is the spectral radius of the adjacency matrix [4, 5]. This is known to represent a lower bound of the real epidemic threshold, yet generally accurate [39] as any correction to it is suppressed by the inverse of the size of the network [43].

In order to extend this approach to temporal networks we need to take into account the time dependence of $\mathbf{\Lambda}$. The Markov chain equations of the process read in this case:

$$p_{t,i} = 1 - \left[1 - (1 - \mu)\, p_{t-1,i}\right] \prod_j \left(1 - \Lambda_{t-1,ji} p_{t-1,j}\right). \qquad (3)$$

We enforce the existence of the asymptotic solution of the infection process in a generic temporal network by imposing periodic boundary conditions for network dynamics, i.e. $\mathbf{W}_{T+1} \equiv \mathbf{W}_1$. Given that T is arbitrary, this causes no loss in generality. We also tested that it would affect the epidemic threshold estimation only for rather small values of T, also when complex temporal dynamics are considered [30]. Contrary to the static case, now the asymptotic solutions of Eq. (3) are periodic of period T.

We develop the formalism introduced in [30], and compute the epidemic threshold for a generic weighted directed temporal network. We define a new representation of the SIS dynamics on a temporal network by employing a multi-layer representation [31, 32, 44]. We map the temporal network to the tensor space $\mathbb{R}^N \otimes \mathbb{R}^T$, where each node is identified by the pair of indices (i, t), corresponding to the node label i and the time frame t respectively. The specific multi-layer representation of the temporal network is built according to the following rules [30]:

(i) each node, at time t, is connected to its future self-image at $t + 1$;

(ii) if i is connected to j at time t with weight w, then we connect i at time t to j at time $t + 1$, and j at time t to i at time $t + 1$, both with weight w.

The rules above define a tensor representation of a weighted multilayer network [32, 45]. We stress that no links connect nodes on the same layer, since layers cease to correspond to the adjacency matrices of the temporal snapshots. The resulting network is thus multipartite, since only pairs of nodes belonging to different layers are linked together. The adjacency representation of the resulting multilayer network has as entries $\hat{W}_{tt',ij} = \delta_{t,t'+1} \left[\delta_{ij} + W_{t,ij}\right]$. The proposed mapping from the network temporal sequence to a multilayer object provides an *ad hoc* representation of the temporal network that preserves the causality of the temporal network and that it lends itself to the integration of the infection and recovery processes. The transformation for the links (rule (ii)) is similar to the one introduced in [46], and is here introduced to model the infection process along a time-stamped link. In addition, we also need to consider the connection between each node and its future self (rule (i)) to model the recovery process of each infected node. We can therefore define the transmission tensor $\mathbf{M}$, whose entries are defined as:

$$M_{tt',ij} = \delta_{t,t'+1} \left[(1 - \mu)\, \delta_{ij} + \Lambda_{t,ij}\right]. \qquad (4)$$

$\mathbf{M}$ contains the transmission terms $\mathbf{\Lambda}_t$ and the recovery term. $\mathbf{M}$ also introduces a simplified expression for Eq. (3). Using the supra-adjacency matrix formalism [31, 47, 48], we can flatten out the multilayer representation using the following mapping: $(i, t) \rightarrow \alpha = Nt + i$, with α running in $\{1, ..., NT\}$, allowing us to write M in matrix form

$$\mathbf{M} = \begin{pmatrix} 0 & 1-\mu+\mathbf{\Lambda}_1 & 0 & \cdots & 0 \\ 0 & 0 & 1-\mu+\mathbf{\Lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1-\mu+\mathbf{\Lambda}_{T-1} \\ 1-\mu+\mathbf{\Lambda}_T & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

M provides a representation of the topological and temporal dimensions underlying the dynamics of Eq. (3), in terms of a $NT \times NT$ transmission matrix that encodes both pathogen transmission and recovery. The Markov process is now represented in $\mathbb{R}^{NT}$ by the state vector $\hat{p}_\alpha(\tau)$, i.e. the probability of each node to be infectious at each time step t included in a 1-period long interval, $[\tau T, (\tau + 1)T]$. Consistently, Eq. (3) becomes

$$\hat{p}_\alpha(\tau) = 1 - \prod_\beta \left[1 - M_{\beta\alpha}\hat{p}_\beta(\tau - 1)\right]. \tag{5}$$

Given that vector $\hat{p}$ encodes a 1-period configuration, the T-periodic asymptotic state of the SIS process is now mapped into the steady state $\hat{p}_\alpha(\tau) = \hat{p}_\alpha(\tau - 1)$. The latter can be recovered as solution of the equilibrium equation:

$$\hat{p}_\alpha = 1 - \prod_\beta \left(1 - M_{\beta\alpha}\hat{p}_\beta\right), \tag{6}$$

that is formally the same as the stationary condition imposed on Eq. (2) for the static network case and is similar to Markov chain approaches used to solve contagion processes in multiplex and interconnected networks [47–49]. Given that Eq. (6) formally describes a diffusion process on a static network of NT nodes, we can then follow [4,5] and linearize Eq. (6) recovering the necessary and sufficient condition for the asymptotically stable disease-free solution, $\rho[\mathbf{M}] < 1$ [50]. This yields the threshold condition

$$\rho[\mathbf{M}] = 1 \tag{7}$$

for the critical value of λ above which the transmission becomes epidemic [4, 5, 47–49]. Given the block structure of the matrix, it is possible to simplify the computation of the spectral radius of $\mathbf{M}$ [51]:

$$\rho[\mathbf{M}] = \rho[\mathbf{P}]^{1/T} \tag{8}$$

where

$$\mathbf{P} = \prod_{t=1}^{T} (1 - \mu + \mathbf{\Lambda}_t). \tag{9}$$

In the case of unweighted undirected network, P becomes $\mathbf{P} = \prod_{t=1}^{T} (1 - \mu + \lambda \mathbf{A}_t)$ [30]. This matrix has an important physical interpretation. Let us consider a time-respecting path from i to j, lasting T time steps and containing a jumps and $T - a$ waiting times. We associate to this path the weight $\lambda^a(1 - \mu)^{T-a}$, representing the probability that the infection propagates along that path, from i infectious at time $t = 1$ to j infectious at time $t = T$. The entry P_{ij} is then the sum of all the time-respecting paths going from $(i, t = 1)$ to $(j, t = T)$, each weighted as described. Therefore, it represents the total probability of j being infectious at time $t = T$, given that the infection originated in i infectious at time $t = 1$. This is valid in the limit of small probabilities and non-interaction among paths. $\mathbf{P}$ thus describes the infection propagation around the disease free state (i.e. $p \simeq 0$) and within the quenched mean field framework where interactions among paths are disregarded. In light of this interpretation, we call $\mathbf{P}$ *infection propagator*. The accessibility matrix, defined in [52], is a particular case of infection propagator, when $\lambda = \mu = 1$, i.e., when the spreading process is a deterministic exploration of the temporal network. The generalization to weighted network is straightforward once the force of transmission on each link of Eq. (1) is taken into account.

3 Infection propagator approach for SIRS and SIR dynamics

Many pathogens leave recovered individuals immune to reinfection. Such immunity may last indefinitely, or for a limited amount of time and is modeled through the introduction of an additional compartment, the recovered (R) state [1, 2]. Infectious nodes enter the recovered state with probability μ, becoming immune to re-infection. We also consider that they leave this state with probability ω, returning to the susceptible state. Any value of $\omega > 0$ describes a SIRS model, characterized by an average immunity period $1/\omega$. $\omega = 0$ corresponds instead to the SIR model, where immunity is assumed to be permanent. Markov chain equations for the SIRS model are as follows:

$$\begin{cases} p_{t,i} = (1 - \mu)\, p_{t-1,i} + \\ \quad + (1 - p_{t-1,i} - q_{t-1,i}) \left[1 - \prod_j \left(1 - \Lambda_{t-1,ji} p_{t-1,j} \right) \right] ; \\ q_{t,i} = \mu p_{t-1,i} + (1 - \omega) q_{t-1,i}. \end{cases} \qquad (10)$$

In addition to $p_{t,i}$, we define $q_{i,t}$ as the probability of being in the recovered state at time t. The computation of the threshold is equivalent to the study of the stability of the disease-free state $p_{t,i} = 0$ [4, 5]. Equations are therefore linearized around that point, making all quadratic terms disappear. In the case of the SIRS model, the disease-free state is $p_{t,i} = q_{t,i} = 0$. Once we linearize Eq. (10) in both $p_{t,i}$ and $q_{t,i}$ we obtain $p_{t,i} = \sum_j \left(\Lambda_{t-1,ji} + (1 - \mu)\delta_{ij} \right) p_{t-1,j}$. The equation for $\mathbf{p}_t$ therefore decouples from $\mathbf{q}_t$, suggesting that the recovered compartment does not impact the epidemic threshold. As a result, the same infection propagator describing the SIS dynamics (Eq. 9) can be written to compute the epidemic threshold of a SIRS compartmental model on a time-varying network.

The SIR model can be considered as a limiting case of the SIRS dynamics ($\omega \to 0$). The infection propagator for the SIRS model does not depend on the probability of waning of immunity ω, as it only contains expressions in terms of λ and μ. The threshold computed with the infection propagator approach for the SIRS therefore holds for any arbitrarily small ω. As a result, we can safely perform the following limit:

$$\lambda_{critical}^{SIR} = \lim_{\omega \to 0} \lambda_{critical}^{SIRS} = \lambda_{critical}^{SIRS} = \lambda_{critical}^{SIS}. \qquad (11)$$

Both SIRS and SIR models thus have the same threshold as the SIS compartmental model, not being affected by the recovery compartment and the duration of the immunity period.

4 Application to empirical and synthetic model data

We test the validity and accuracy of our predictions by comparing them with the results of explicit microscopic numerical simulations of the SIR and SIRS processes. We consider two temporal network models and one empirical time-varying network. In the following subsections we describe the data and methods considered and the corresponding results.

4.1 Empirical and synthetic model data

We test our approach on two network models: the ACTIVITY and BURSTY models. ACTIVITY is built from the activity-driven model proposed by Perra *et al.* in [25]. Each node is given an activity potential, drawn from a heterogeneous distribution. At each time step, nodes become active with a probability equal to their potentials. Active nodes establish m (here $m = 2$) connections with other nodes picked at random, and all links are renewed at every snapshot. We generate networks with $N = 1000$ nodes and $T = 20$ time snapshots, as in [30]. Activity potentials are assigned through the relation $a = 1 - e^{-\eta x}$, ($\eta = 10$) and $x \sim x^{-\gamma}$ and $x \in [\epsilon, 1]$ ($\gamma = 2.8$ and $\epsilon = 3 \cdot 10^{-2}$). The obtained networks are characterized by a temporally uncorrelated sequence of snapshots displaying a heterogeneous topology.

BURSTY is obtained from the model introduced by Rocha *et al.* in [20]. Here, the probability of a node becoming active at time step t is sampled from the distribution $(t - t')^{-\alpha_1} e^{-\alpha_2(t-t')}$, where t' is the time that node was last active. We consider networks of size $N = 500$ and described by $T = 50$ time snapshots, generated with $\alpha_1 = 2$ and $\alpha_2 = 5 \cdot 10^{-4}$ as in [30]. The obtained networks account for a heterogeneous activation pattern describing a sequence of homogeneous networks where the inter-contact time is power-law distributed

In addition to the synthetic models above, we consider an empirical time-evolving network constructed from records of face-to-face proximity interactions between individuals in a high school during one day, collected by Salathé *et al.* [53] (SCHOOL). This network comprises $N = 787$ nodes and we consider here $T = 42$ time snapshots, each one of 10 minutes.

4.2 Numerical simulations

We numerically simulate the disease diffusion of a SIR and of a SIRS infection dynamics on the above described networks. Simulations assume all individuals to be susceptible at the initial time, and are seeded with an infected node chosen at random on the network. At each time step, infectious nodes can transmit the disease with probability λ to their susceptible neighbors and recover with probability μ. Here we consider unweighted networks, for the sake of simplicity. Weighted networks will be addressed in the next section in the study of time aggregation of the evolving network. In the SIRS model, recovered nodes turn susceptible with probability ω. Results of the simulations are obtained after randomizing the initial seed and the time step of the T sequence chosen as the initial time step, and they are obtained under the assumption of periodic boundary conditions for network evolution.

For the SIRS dynamics, following [54] we numerically identify the epidemic threshold as the value of the transmissibility λ for which the relative variation of the prevalence at equilibrium is maximal, as such variation would go to infinity in the thermodynamic limit ($N \to \infty$), indicating a second order phase transition. We therefore measure the *variablity* $\Delta = \sqrt{\langle i_T^2 \rangle - \langle i_T \rangle^2} / \langle i_T \rangle$ [54–56], where $i_T = \frac{1}{T} \sum_{t=1}^{T} i_{eq}(t)$ is the prevalence at equilibrium averaged over a period T. The endemic prevalence is computed using the quasistationary method [54, 57]. We force the system to be in an active state; whenever it reaches the absorbing state with no infectious nodes, we sample one random configuration

among the ones the system had visited while it was in the same snapshot, and restart the simulation from that configuration. After discarding an initial transient ($3 \cdot 10^3$ iterations), we compute i_T for every period (for $5 \cdot 10^5$ iterations), and with those values we compute $\langle i_T \rangle$ and $\langle i_T^2 \rangle$. We then compare the value of λ corresponding to the peak of the variability Δ with the prediction for the epidemic threshold obtained from the infection propagator approach. The same method is used for the SIR dynamics, where the endemic prevalence is replaced by the final attack rate r, i.e. the fraction of nodes hit by the epidemic

4.3   Results

We consider a SIR dynamics on the three networks under study and explore two values of the infectious period, corresponding to $\mu = 0.2, 0.5$. For each μ, Figure 1 shows the behavior of the variability Δ normalized to its peak value Δ_{max} as a function of the transmissibility λ. In all cases we find a very good agreement between our prediction (vertical dashed line) and the simulated epidemic threshold obtained from the peak value of Δ. The agreement is found for both network models ACTIVITY and BURSTY, despite them being characterized by different topological and temporal heterogeneities, and for the empirical dataset SCHOOL. This last network features a more complex dynamics capturing the daily activities and interactions, with non-trivial temporal correlations and modular structures evolving in time [53]. Despite the approximations used to compute the epidemic threshold with the infection propagator approach, the results of Figure 1 indicate that the method is able to provide reliable and accurate predictions for the threshold behavior of systems characterized by different properties. We also note that the agreement is obtained independently of the values of the epidemic threshold: the threshold of the SCHOOL network is indeed approximately one order of magnitude smaller than the ones obtained in the two network models for the same SIR dynamics.

Similar results are also obtained when considering a SIRS dynamics, characterized by the same values of the infectious period considered above and by three values of the probability of immunity waning ($\omega = 0.25, 0.5, 0.75$). For each temporal network, we numerically identify the value of the epidemic threshold as that corresponding to the peak of the normalized variability Δ/Δ_{max}, and recover a good agreement with our analytical predictions (Figure 2). The addition of the transition from an immune state to a susceptible state does not alter the accuracy of the computed predictions. Moreover, different immunity periods (i.e. different values of ω) lead to the same epidemic threshold on the temporal networks, as predicted by the infection propagator approach. The difference observed in the curves for different values of ω for λ well above the threshold is induced by the variation in the average endemic prevalence. Epidemics circulating on these systems and characterized by longer immunity periods ($\omega = 0.25$, light blue and light red in the plots of Figure 2) display a larger variability due to the smaller average prevalence reached at equilibrium, as shown by Figure 3. Shorter immunity periods ($\omega = 0.75$) reach a larger endemic prevalence for a given value of the transmissibility above the epidemic threshold and therefore display a smaller variability Δ/Δ_{max}.
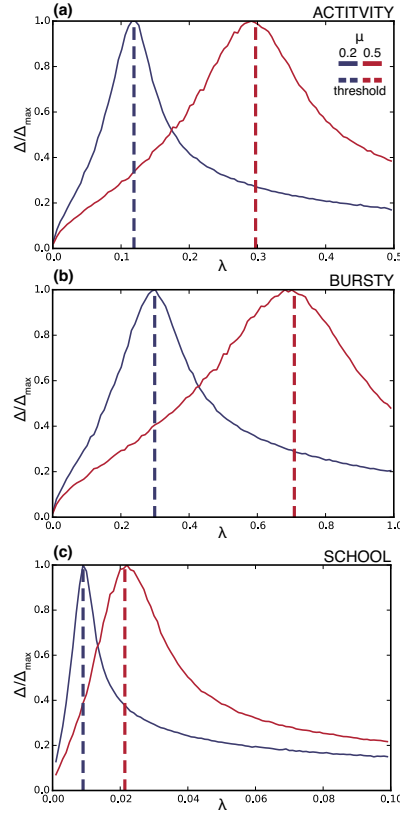
Figure 1: **SIR model. Comparison between the epidemic threshold and variability.** The variability Δ of final attack rate r, normalized to its peak value (Δ/Δ_{max}), is plotted against the transmissibility λ. We explore two different values of μ. Dashed vertical lines represent the threshold value for λ predicted by the infection propagator approach. (a) shows the results for the ACTIVITY network, (b) for the BURSTY network, and (c) for the SCHOOL network.

5   Impact of time aggregation of the temporal network

In many cases, information on network dynamics can be coarse, with data reporting on the temporal evolution at a lower resolution scale than the one of the process itself. This means that all events occurring within the time interval of the considered resolution will be aggregated in a static single snapshot. An aggregated representation of a temporal network does not account for causal structures and temporal correlations that occur at time scales that are smaller than aggregation interval [58]. Since these structures can impact disease dynamics, it is crucial to assess how such coarser representation influences the description of epidemic processes [37, 58, 59]. Here, we study the influence of the aggregation schemes described in [59] on the epidemic threshold. HET scheme is a weighted aggregation of the snapshots, obtained by summing link weights: $\mathbf{W}_t, \mathbf{W}_{t+1} \longmapsto \mathbf{W}_t + \mathbf{W}_{t+1}$. By contrast, HOM is topologically equivalent to HET, having the exact same set of links. However, each link is given an equal weight corresponding to the average weight of the HET network over the same period. As a result, both schemes
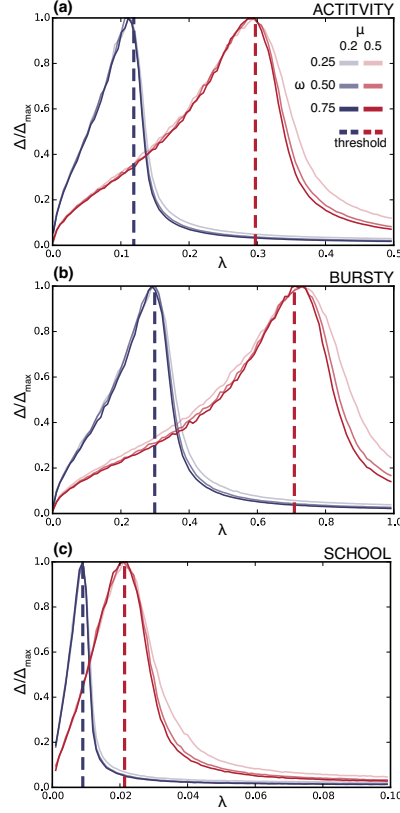
Figure 2: **SIRS model. Comparison between the epidemic threshold and variability.** The variability Δ of the endemic prevalence i_T, normalized to its peak value (Δ/Δ_{max}), is plotted against the transmissibility λ. We explore different values of μ and ω. Dashed vertical lines represent the threshold value for λ predicted by the infection propagator approach. (a) shows the results for the ACTIVITY network, (b) for the BURSTY network, and (c) for the SCHOOL network.

share the same average weight at every aggregation interval, but HET accounts for weight heterogeneity. We use the intrinsic transmissibility λ for comparison across different aggregation schemes and intervals, as it does not depend on weight. We consider the empirical dataset of the SCHOOL network as it provides a richer temporal and topological set of features with respect to synthetic models. Also, the study on aggregation aims at providing useful practical information for data collection purposes.

We consider the highest resolution network obtained from the SCHOOL data corresponding to $\Delta t_1 = 20s$. Starting from this resolution, we aggregate snapshots recursively two by two, doubling the aggregation interval at each aggregation test. We consider the recovery rate m as an intrinsic property of the disease, thus not changing with aggregation. The probability of recovery after a time Δt is $me^{-m\Delta t}$. Aggregation interval at the k-th aggregation is $\Delta t_k = k\Delta t_1$. Hence, we compute the recovery probability at the k-th aggregation $\mu[\Delta t_k]$ as the probability of recovering within an interval Δt_k, i.e., $\mu[\Delta t_k] = 1 - e^{-m\Delta t_1 k}$. We explore four different recovery rates: $m = 1.8, 9, 18, 90$ h^{-1}.
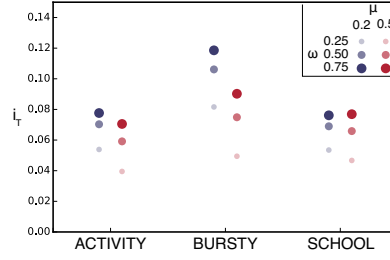
Figure 3: **SIRS model. Comparing the average prevalence above threshold,
for different values of ω.** For each network we choose a value of λ
above threshold, so that variability is around $1/3$ of its peak, $\Delta/\Delta_{max} \approx$
$1/3$. For ACT this corresponds to $\lambda = 0.14$ for $\mu = 0.2$, and $\lambda = 0.35$ for
$\mu = 0.5$; for BURSTY $\lambda = 0.36$ and $\lambda = 0.86$, and for SCHOOL $\lambda = 0.01$
and $\lambda = 0.03$. We plot the average prevalence at equilibrium i_T for all
three network, and for all the explored values of μ and ω.

5.1 Results

We compare the epidemic threshold $\lambda_{\Delta t}$ computed after aggregating the network with a
given aggregation time window Δt, to the one computed at the highest resolution Δt_1
(λ_1), using the ratio $\lambda_{\Delta t}/\lambda_1$ (Figure 4). Four values of the recovery rate are considered,
and for each value the results from two time aggregation schemes are shown.

Focusing on the HET aggregation scheme, the results of Figure 4(a) show that the
prediction made on the aggregated SCHOOL network deteriorates with the increase of the
time aggregation window Δt. As expected, the aggregation induces a loss of the temporal
information making the aggregated network to perform poorly with respect to reproduc-
ing the behavior obtained in the original network. This is known for a series of indicators
regarding the importance of individual nodes in the spread of an epidemic in the sys-
tem [58], and we find that it also results in a biased estimation of the threshold condition
for the epidemic propagation. The effect is more rapid and stronger for fast epidemic
processes (e.g. $m = 90$ hours^{-1}), in that a fast disease circulating on the network would
have the possibility to experience the entire landscape of dynamical changes the network
undergoes through, thus differentiating between the pattern obtained at the highest reso-
lution and the aggregated one. On the other hand, if the disease spreads slowly on the
network (e.g. $m = 1.8$ h^{-1}), we expect the epidemic process to be less sensitive to the
network changes. The epidemic threshold computed on the aggregated network provides
indeed a good estimate of the one corresponding to the highest resolution network up to
a certain level of aggregation (e.g., $\Delta t \simeq 3min20s$ for the $m = 1.8$ hours^{-1} case), after
which the accuracy is progressively lost. This is consistent with the numerical results of
an SEIR dynamics spreading on the network of contact of conference attendees, show-
ing that the spreading dynamics is well described by a static aggregated network if the
heterogeneity of the contact durations is taken into account as edge weights [59].

The underlying mechanism leading to the deterioration of the epidemic threshold esti-
mate with increasing aggregation time window is the creation of novel transmission paths
that would otherwise not exist, with the effect of destroying the causality of the sequence
of interactions within Δt and of increasing the density of the links in the network [60,61].

All these effects tend to facilitate the spread of a disease, so that the resulting epidemic threshold is lower than the one computed on the original temporal network corresponding to $\Delta t = 1$, as shown in Figure 4(a).

If we focus on the HOM aggregation scheme, we observe that the epidemic threshold predicted for a given Δt is systematically higher than the one obtained in the HET scheme for the same Δt value (same color, dashed lines vs. continuous lines in Figure 4(a)). The reason lies in the way weights are distributed over the links of the aggregated networks. While the HET scheme preserves the heterogeneity of the duration of the contacts, cumulating the duration of the interaction established by each pair of individuals, this information is lost in the HOM scheme as the total contact duration is homogeneously distributed among all contacts. Heterogeneity of the weights has a strong effect on the evolution of epidemics [45, 62–70], favoring the spread of diseases [21, 71–73]. This results in a lower epidemic threshold than its homogeneous counterpart, for a given Δt. The faster the disease is, the smaller is the difference observed in the epidemic threshold obtained from the two aggregation schemes.

To better explore the various facets of the SCHOOL temporal network having an impact on the threshold condition, we also consider three reference models that systematically destroy some of the network properties. RESHUFFLE consists of a random reshuffling of snapshot time ordering. It preserves the aggregated network, and the static topological features of the snapshots. It breaks the temporal activity of the network, defined as the number of contacts in time. It breaks all temporal correlations among link activations, too. RECONFIGURE consists of a random reassignment of contact timestamps. Two contacts (i, j, t), (k, l, s) are randomly selected, and their timestamp switched: (i, j, s), (k, l, t). RECONFIGURE preserves the activity timeline and the aggregated network. It breaks snapshot topology and temporal correlations between link activations. Finally, ANONYMIZE reshuffles the identity of the nodes of each time snapshot, thus preserving activity timeline and static topology of each snapshot. It breaks all dynamic community structures and cliques (namely school classes).

Results for these reference models are shown in Figure 4(b)-(d). The behaviors obtained for RESHUFFLE and RECONFIGURE models are very similar. The difference between HET and HOM schemes is reduced for all recovery rates with respect to the results obtained on the original network, and it becomes negligible for faster diseases. The curves of panels (b) and (c) show that the obtained result is independent of the activity timeline of the network (preserved by the RECONFIGURE reference model, but not by the RESHUFFLE one), and it is more likely related to specific time-evolving topological structures present in the SCHOOL network that are otherwise destroyed by both reference models. To test this hypothesis, we consider the ANONYMIZE reference model (Figure 4(d)), where we destroy all two-points correlations and their time correlations, while preserving the overall temporal activity and the topology of each snapshot. As expected, the two schemes cannot be anymore distinguished following such reshuffling.

Results of Figure 4 show that in all reference models aggregation leads to an underestimation of the epidemic threshold, for both aggregation schemes considered. In the SCHOOL network, on the other hand, HOM aggregation is found to provide larger epidemic thresholds than the one obtained at the highest resolution, within a given aggregation interval and for slow diseases. To better understand this behavior observed solely on the empirical data that disappears with the three types of reshuffling considered in the refer-
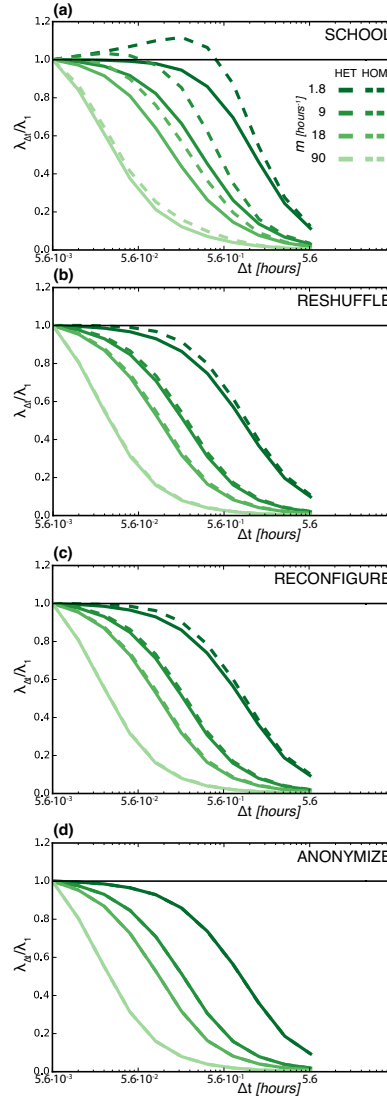
Figure 4: **Impact of aggregation on the epidemic threshold for the** SCHOOL **network and three reference models.** The ratio between the threshold $\lambda_{\Delta t}$ computed on the aggregated network and the threshold λ_1 of the highest resolution temporal network is plotted as a function of the aggregation time interval Δt. Four different values of the recovery rates are explored, along with two aggregation schemes, HOM and HET. (a) shows the results for SCHOOL, (b) for reference model RESHUFFLE, (c) for RECONFIGURE, and (d) for ANONYMIZE.

ence models, we explore the role of time correlations and memory effects in the SCHOOL network. We consider the *social strategy* introduced in [74]. More in detail, we fix a time window of $\delta = 20$ snapshots ($6min 40s$), and define $k_{t,i}^{HET}$ ($k_{t,i}^{HOM}$) as the degree of node i in the network aggregated over the interval $[t-\delta, t]$, using aggregation scheme HET (HOM). We compute the social strategy of node i at time step t as $\gamma_{t,i} = k_{t,i}^{HOM}/k_{t,i}^{HET}$ (the same definition as in [74], except for a normalizing factor δ). Social strategy discriminates between memory-driven behavior ($\gamma \to 0$), where a node tends to make contacts always

with the same nodes, and memoryless behavior ($\gamma \to 1$), where a node shows a more socially exploratory behavior. Figure 5(a) shows how social strategy evolves in time. We observe that its median behavior is quite stable in time, except for several localized spikes. Most of these spikes roughly correspond to abrupt variations in the temporal activity of the network. These spikes result then from a reduction of the memory of the system, due to a varying number of overall contacts. Remarkably, however, the median social strategy returns to the value it had before the spike quickly after each of these events, indicating that the interaction dynamics does not qualitatively change, but the sets of interacting individuals do change over time. The only exception occurs between around $13 : 30$ and $14 : 30$, when social strategy is significantly higher than average, but still lower than its delimiting peaks. These spikes naturally induce a temporal slicing of the network, in a way that likely corresponds to the rhythm of school activities. We call γ-*slice* each time interval between two consecutive spikes.

This aspect of the degree of memory contained in the system, and measured through the tendency of each node to keep establishing contacts with the same individuals over time, is destroyed in all reference models under study, even those that preserve the activity timeline of Figure 5(a). To understand whether if and to what extent stages in the evolution of the social strategy in time are responsible for the behavior observed in the SCHOOL network, we design a fourth reference model, RESHUFFLE-SOCIAL, where we randomize the snapshot order, as in RESHUFFLE, but we allow reshuffling only within each γ-slice. Figure 5(b) shows that RESHUFFLE-SOCIAL displays the same behavior as the SCHOOL network, unlike RESHUFFLE, with an overestimation of the value of the epidemic threshold by the HOM scheme for small enough aggregation intervals and slow diseases. The aggregation of snapshots where individuals show a rather large memory in the way they establish links (i.e. small γ) leads to marked weight-topology correlations, likely being part of robust temporal communities of highly interacting nodes emerging from school daily activities. Such correlations were already found to play an important role in the slowing down of epidemics once large-scale propagation occurs in the system [18]. In our case, we find that preserving the heterogeneity of weights of such correlations (as in the HET scheme) can provide a good approximation of the epidemic threshold for small interval aggregation and for slow diseases. In addition, such approximation is better than the one provided by homogenizing weights across all links in the system (as in the HOM scheme), given that the latter destroys weight-topology correlations leading to a network that is more resilient to the epidemic spread [21]. This effect vanishes for increasing time aggregating windows and it completely disappears for all aggregating intervals once these correlations are destroyed by the reshuffling of nodes (see the ANONYMIZE reference model in Figure 4).

6 Conclusions

We have considered the infection propagator approach to compute the epidemic threshold for an arbitrary time-varying network. Starting from a SIS dynamics on a weighted directed temporal network, we have considered more complicated compartmental models and addressed timescale issues relevant for the study of temporal networks. The overall aim was to introduce the infection propagator approach for more realistic infection dynamics and to study the effect of time aggregation of the network of contacts on the
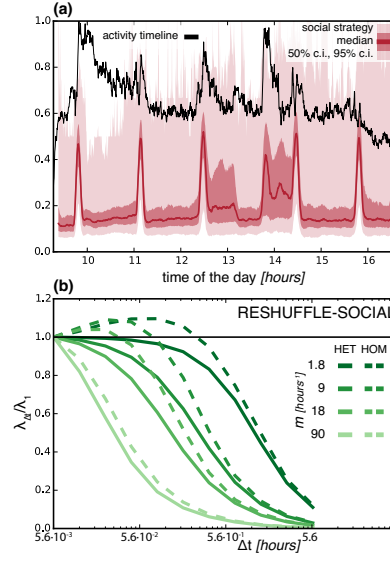
Figure 5: **Interplay between** SCHOOL **dynamics and aggregation.** In (a) we show the temporal activity of SCHOOL, as the normalized number of contacts (black line) in each snapshot of the fully temporal network. On the x axis we indicate the time of the day, in hours. Red solid line represents the median value of network's social strategy, computed with a sliding window of 20 snapshots (equivalent to $400s$). Social strategy plotted at time t is computed over the interval $(t-20, t]$. Red areas show 50% (darker) and 95% (lighter) confidence interval for social strategy. In (b) we plot, for reference model RESHUFFLE-SOCIAL, the ratio between the threshold computed on the aggregated network $\lambda_{\Delta t}$, and the threshold of the full temporal network λ_1, as a function of the aggregation time interval Δt. Four different values of the recovery rates are explored, along with the two aggregation schemes HOM and HET.

computation of its threshold. Our findings indicate that the approach provides reliable and accurate predictions of the epidemic threshold also in presence of immunity stages and loss of immunity transitions in the disease natural history. In addition, for slow diseases, the time aggregation scheme preserving the cumulative heterogeneous duration of contacts between two nodes is shown to provide a quite accurate estimation of the epidemic threshold of the corresponding high-resolution network up to a certain aggregation level. For faster diseases, time aggregation strongly alters the accuracy of the estimation. The presence of weight-topology correlations is the main feature of the SCHOOL network leading to biased estimations. These findings provide important information to study the vulnerability of systems in real settings and to assess possible biases induced by the consideration of time-aggregated contact data.

References

[1] R M Anderson and R M May. *Infectious Diseases of Humans Dynamics and Control*. Oxford University Press, 1992.

[2] Matthew James Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton Univ Pr, 2008.

[3] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14):3200–3203, April 2001.

[4] Yang Wang, D. Chakrabarti, Chenxi Wang, and C. Faloutsos. Epidemic spreading in real networks: an eigenvalue viewpoint. In *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, pages 25–34, Oct 2003.

[5] Sergio Gómez, Alexandre Arenas, J Borge-Holthoefer, Sandro Meloni, and Yamir Moreno. Discrete time Markov chain approach to contact-based disease spreading in complex networks. *Europhys. Lett.*, 2010.

[6] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626–4628, Nov 2000.

[7] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, Jul 2002.

[8] Marian Boguñá, Claudio Castellano, and Romualdo Pastor-Satorras. Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks. *Phys. Rev. Lett.*, 111:068701, Aug 2013.

[9] Claudio Castellano and Romualdo Pastor-Satorras. Thresholds for Epidemic Spreading in Networks. *Phys. Rev. Lett.*, 105:218701, November 2010.

[10] A. V. Goltsev, S. N. Dorogovtsev, J. G. Oliveira, and J. F. F. Mendes. Localization and spreading of diseases in complex networks. *Phys. Rev. Lett.*, 109:128702, Sep 2012.

[11] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[12] S Bansal, J Read, and B Pourbohloul. The dynamic nature of contact networks in infectious disease epidemiology. *Journal of Biological Dynamics*, 2010.

[13] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1):166–180, February 2011.

[14] Giovanna Miritello, Esteban Moro, and Rubén Lara. Dynamical strength of social ties in information spreading. *Phys. Rev. E*, 83:045102, Apr 2011.

[15] Luis E C Rocha, Fredrik Liljeros, and Petter Holme. Information dynamics shape the sexual networks of Internet-mediated prostitution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13):5706–5711, March 2010.

[16] Paolo Bajardi, Alain Barrat, Fabrizio Natale, Lara Savini, and Vittoria Colizza. Dynamical patterns of cattle trade movements. *PloS ONE*, 6(5):e19869, 2011.

[17] Carter T. Butts. Revisiting the foundations of network analysis. *Science*, 325(5939):414–416, 2009.

[18] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E*, 83:025102, Feb 2011.

[19] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 2012.

[20] Luis E. C. Rocha and Vincent D. Blondel. Bursts of vertex activation and epidemics in evolving networks. *PLoS Comput Biol*, 9(3):e1002974, 03 2013.

[21] Luca Ferreri, Paolo Bajardi, Mario Giacobini, Silvia Perazzo, and Ezio Venturino. Interplay of network dynamics and heterogeneity of ties on spreading dynamics. *Phys. Rev. E*, 90:012812, Jul 2014.

[22] José Luis Iribarren and Esteban Moro. Impact of human activity patterns on the dynamics of information diffusion. *Phys. Rev. Lett.*, 103:038702, Jul 2009.

[23] Alexei Vazquez, Balázs Rácz, András Lukács, and Albert-László Barabási. Impact of non-poissonian activity patterns on spreading processes. *Phys. Rev. Lett.*, 98:158702, Apr 2007.

[24] Ken T.D. Eames and Matt J. Keeling. Monogamous networks and the spread of sexually transmitted diseases. *Mathematical Biosciences*, 189(2):115 – 130, 2004.

[25] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Scientific Reports*, 2, June 2012.

[26] Erik Volz and Lauren Ancel Meyers. Epidemic thresholds in dynamic contact networks. *Journal of The Royal Society Interface*, 6(32):233–241, 2009.

[27] Thilo Gross, Carlos J. Dommar D'Lima, and Bernd Blasius. Epidemic dynamics on an adaptive network. *Phys. Rev. Lett.*, 96:208701, May 2006.

[28] Michael Taylor, Timothy J. Taylor, and Istvan Z. Kiss. Epidemic threshold and control in a dynamic network. *Phys. Rev. E*, 85:016103, Jan 2012.

[29] Zhenyuan Zhao, J P Calderón, Chen Xu, Guannan Zhao, Dan Fenn, Didier Sornette, Riley Crane, Pak Ming Hui, and Neil F Johnson. Effect of social group dynamics on contagion. *Phys. Rev. E*, 81(5):56107, May 2010.

[30] Eugenio Valdano, Luca Ferreri, Chiara Poletto, and Vittoria Colizza. Analytical Computation of the Epidemic Threshold on Temporal Networks. *Phys. Rev. X*, 5(2):21005, April 2015.

[31] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A. Porter, Sergio Gómez, and Alex Arenas. Mathematical formulation of multilayer networks. *Phys. Rev. X*, 3:041022, Dec 2013.

[32] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer Networks. *Journal of Complex Networks*, 2014.

[33] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, and Nicolas Voirin. Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. *PLoS ONE*, 8(9):e73970, 2013.

[34] Mario Konschake, Hartmut H K Lentz, Franz J Conraths, Philipp Hövel, and Thomas Selhorst. On the Robustness of In- and Out-Components in a Temporal Network. *PLoS ONE*, 8(2):e55223, 2013.

[35] Julie Fournet and Alain Barrat. Contact Patterns among High School Students. *PLoS ONE*, 9(9):e107878, 2014.

[36] Thomas Obadia, Romain Silhol, Lulla Opatowski, Laura Temime, Judith Legrand, Anne C M Thiébaut, Jean-Louis Herrmann, Éric Fleury, Didier Guillemot, and Pierre-Yves Boëlle. Detailed Contact Data and the Dissemination of Staphylococcus aureus in Hospitals. *PLoS Comput Biol*, 11(3):e1004170, 2015.

[37] Bruno Ribeiro, Nicola Perra, and Andrea Baronchelli. Quantifying the effect of temporal resolution on time-varying networks. *Sci. Rep.*, 3, October 2013.

[38] Paolo Bajardi, Alain Barrat, Lara Savini, and Vittoria Colizza. Optimizing surveillance for livestock disease spreading through animal movements. *Journal of The Royal Society Interface*, June 2012.

[39] James P. Gleeson, Sergey Melnik, Jonathan A. Ward, Mason A. Porter, and Peter J. Mucha. Accuracy of mean-field theory for dynamics on real-world networks. *Phys. Rev. E*, 85:026106, Feb 2012.

[40] Piet Van Mieghem. The n-intertwined sis epidemic network model. *Computing*, 93(2-4):147–169, December 2011.

[41] P. Van Mieghem and R. van de Bovenkamp. Non-markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks. *Phys. Rev. Lett.*, 110:108701, Mar 2013.

[42] Cong Li, Huijuan Wang, and Piet Van Mieghem. Epidemic threshold in directed networks. *Phys. Rev. E*, 88:062802, Dec 2013.

[43] E. Cator and P. Van Mieghem. Second-order mean-field susceptible-infected-susceptible epidemic threshold. *Phys. Rev. E*, 85:056111, May 2012.

[44] Wehmuth Klaus, Ziviani Artur, and Fleury Eric. A unifying model for representing time-varying graphs. *arXiv.org*, February 2014.

[45] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 329:277, 2010.

[46] Ingo Scholtes, Nicolas Wider, René Pfitzner, Antonios Garas, Claudio J Tessone, and Frank Schweitzer. Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. *Nature Communications*, 5:5024, 2014.

[47] Emanuele Cozzo, Raquel A. Baños, Sandro Meloni, and Yamir Moreno. Contact-based social contagion in multiplex networks. *Phys. Rev. E*, 88:050801, Nov 2013.

[48] Huijuan Wang, Qian Li, Gregorio D'Agostino, Shlomo Havlin, H. Eugene Stanley, and Piet Van Mieghem. Effect of the interconnected network structure on the epidemic threshold. *Phys. Rev. E*, 88:022801, Aug 2013.

[49] Clara Granell, Sergio Gómez, and Alex Arenas. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Phys. Rev. Lett.*, 111:128701, Sep 2013.

[50] Saber Elaydi. *An introduction to difference equations*. Springer, New York, NY, USA, third edition, 2005.

[51] P D Powell. Calculating Determinants of Block Matrices. *ArXiv e-prints*, 2011.

[52] H K Lentz, T Selhorst, and I M Sokolov. Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks. *Phys. Rev. Lett.*, 110(11), 2013.

[53] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W Feldman, and James H Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 107(51):22020–22025, December 2010.

[54] Silvio C Ferreira, Claudio Castellano, and Romualdo Pastor-Satorras. Epidemic thresholds of the susceptible-infected-susceptible model on networks: a comparison of numerical and theoretical results. *Phys. Rev. E*, 86(4 Pt 1):041125, October 2012.

[55] Pascal Crépey, Fabián P. Alvarez, and Marc Barthélemy. Epidemic variability in complex networks. *Phys. Rev. E*, 73:046131, Apr 2006.

[56] P. Shu, W. Wang, M. Tang, and Y. Do. Simulated identification of epidemic threshold on finite-size networks. *ArXiv e-prints*, October 2014.

[57] Silvio C. Ferreira, Ronan S. Ferreira, Claudio Castellano, and Romualdo Pastor-Satorras. Quasistationary simulations of the contact process on quenched networks. *Phys. Rev. E*, 84:066102, Dec 2011.

[58] P Holme. Epidemiologically optimal static networks from temporal network data. *PLoS Comput Biol*, 9(7):e1003142, 2013.

[59] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Régis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, and Philippe Vanhems. Simulation of an SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine*, 9(1):87, July 2011.

[60] P Holme. Network reachability of real-world contact sequences. *Phys. Rev. E*, 71:046119, 2005.

[61] R. K. Pan and J. Saramaki. Path lengths, correlations, and centrality in temporal networks. *Phys. Rev. E*, 84:016105, 2011.

[62] Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):pp. 1360–1380, 1973.

[63] Yan Gang, Zhou Tao, Wang Jie, Fu Zhong-Qian, and Wang Bing-Hong. Epidemic spread in weighted scale-free networks. *Chinese Physics Letters*, 22(2):510, 2005.

[64] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[65] R. Lambiotte, J.-C. Delvenne, and M. Barahona. Laplacian Dynamics and Multi-scale Modular Structure in Networks. *arXiv.org*, 2009.

[66] R. Toivonen, X. Castelló, V. M. Eguíluz, J. Saramäki, K. Kaski, and M. San Miguel. Broad lifetime distributions for ordering dynamics in complex networks. *Phys. Rev. E*, 79:016109, Jan 2009.

[67] Ken T.D. Eames, Jonathan M. Read, and W. John Edmunds. Epidemic prediction and control in weighted networks. *Epidemics*, 1(1):70 – 76, 2009.

[68] Marcel Salathé and James H Jones. Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Comput Biol*, 6(4):e1000736, 2010.

[69] Zimo Yang and Tao Zhou. Epidemic spreading in weighted networks: An edge-based mean-field solution. *Phys. Rev. E*, 85:056106, May 2012.

[70] Christel Kamp, Mathieu Moslonka-Lefebvre, and Samuel Alizon. Epidemic spread on weighted networks. *PLoS Comput Biol*, 9(12):e1003352, 12 2013.

[71] Maria Deijfen. Epidemics and vaccination on weighted graphs. *Mathematical Biosciences*, 232(1):57 – 65, 2011.

[72] Tom Britton, Maria Deijfen, and Fredrik Liljeros. A weighted configuration model and inhomogeneous epidemics. *Journal of Statistical Physics*, 145(5):1368–1384, 2011.

[73] Tom Britton and David Lindenstrand. Inhomogeneous epidemics on weighted networks. *Mathematical Biosciences*, 240(2):124 – 131, 2012.

[74] Giovanna Miritello, Rubén Lara, Manuel Cebrian, and Esteban Moro. Limited communication capacity unveils strategies for human interaction. *Sci. Rep.*, 3, June 2013.

END

Valdano E, Poletto C, Colizza V

Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution

under review (2015)

3.4. Continuous time

The methodology we have introduced so far intrinsically consider time as a discrete entity, as this allows the multilayer mapping. As we have explained in Section 2.1, this is often a reasonable assumption. There are contexts, however, where one still needs to assume continuous time evolution [87,90]. Hence, in this section we address the problem of computing the threshold on continuous-time varying temporal networks.

We have seen that the infection propagator (Equation 3.2) contains all the information we need. We now develop a calculation to directly compute its continuous limit, in order to obtain the infection propagator in continuous time. Once we have that, we compute the threshold in terms of its spectral radius, as before.

We treat the continuous-time temporal network as an adjacency matrix $A(t)$ which is a continuous function of time $t \in [0, T_c]$. We sample this network at regular intervals, with a sampling interval Δt. We thus get a discrete temporal network $\{A_1, A_2, \cdots, A_T\}$ which is a sample of the original one, with $A_h = A(h\Delta t)$ and $T = T_c/\Delta t$. Through Equation 3.2 we can write the infection propagator for the sampled network. We now wish to recover the propagator of the continuous time network by making Δt go to zero. As we cannot straightforwardly perform such limit on Equation 3.2, we define a partial propagator $P(h)$ as

$$P(h) = \prod_{k=1}^{h} \left[1 - \mu + \lambda A_k \right]. \tag{3.3}$$

This is a truncated version of the infection propagator, reaching the full propagator when $h = T$: $P(T) = P$. One can easily see that this object obeys the recursion relation

$$P(h + 1) = P(h) \left[1 - \mu + \lambda A_{h+1} \right]. \tag{3.4}$$

We exploit the fact that A_h matrices are instantaneous pictures of the continuous process, and rewrite Equation 3.4 as a function of continuous time:

$$P(t + \Delta t) - P(t) = P(t) \left[\mu(\Delta t) + \lambda(\Delta t)A(t + \Delta t) \right]. \tag{3.5}$$

In order to perform the limit $\Delta t \to 0$, we introduce the time-independent transmission and recovery rates l, m. They are intrinsic features of the diseases, and encode the probability of transmission and recovery within a small time interval:

$$\begin{cases} \lambda(\Delta t) = l\Delta t + \mathcal{O}(\Delta t^2); \\ \mu(\Delta t) = m\Delta t + \mathcal{O}(\Delta t^2). \end{cases} \tag{3.6}$$

We rewrite Equation 3.5 translating probabilities into rates, and divide both sides by Δt:

$$\frac{P(t + \Delta t) - P(t)}{\Delta t} = P(t)\left[-m + lA(t + \Delta t)\right] + \mathcal{O}(\Delta t). \tag{3.7}$$

Equation 3.5 is clearly well behaved when Δt becomes very small, and we can finally compute the limit $\Delta t \to 0$. The left hand side becomes, by definition, the derivative of P.

$$\frac{d}{dt}P(t) = P(t)\left[-m + lA(t)\right]. \tag{3.8}$$

If we assume $A(t)$ is a continuous function of time, everything is well behaved. If not, we can perform derivatives in the sense of the distributions [197, 198]. The solution of Equation 3.8 in $t = T_c$, for initial condition $P(0) = 1$, leads to the infection propagator $P = P(T_c)$ in case of continuous time. Equation 3.8, however, cannot be solved explicitly in most cases, since it is a non-autonomous system of coupled differential equations of N^2 unknowns (all entries P_{ij}). It must therefore be solved numerically, but such computation becomes cumbersome, even for moderately large networks. Hence, whenever possible, the discrete approach should be preferred, as it is much more numerically efficient. There is, however, a particular case for which we can explicitly solve Equation 3.8. We must assume that adjacency matrices at different times commute: $[A(t), A(t')] = 0$ $\forall t, t' \in [0, T_c]$. How stringent is this requirement? Let us consider an arbitrary temporal graph, which for simplicity we assume undirected and without self-loops. Its adjacency matrix $A(t)$ then has $N(N - 1)/2$ degrees of freedom, as we can turn on and off each link at any time. We can diagonalize $A(t)$ getting a diagonal matrix $D(t)$, and a matrix of change of basis $Q(t)$, and both will depend on time: $A(t) = Q(t)D(t)Q(t)^{-1}$. Now, imposing commutativity $[A(t), A(t')] = 0$ is equivalent to fixing $Q(t) = Q$ in time. That is because commuting matrices are diagonalized by the same matrix, and so there must exists a single Q that diagonalizes $A(t)$ at any time. Hence, we are down to N degrees of freedom, corresponding to the diagonal entries of $D(t)$. Conceptually, this means that network topology has no temporal correlations or memory, as that would induce non-zero commutator. Let us now solve Equation 3.8 in case of commuting matrices, using $A(t) = QD(t)Q^{-1}$:

$$\frac{d}{dt}P(t) = P(t)Q\left[-m + lD(t)\right]Q^{-1}. \tag{3.9}$$

We define $Y(t) := P(t)Q$, and rewrite

$$\frac{d}{dt}Y(t) = Y(t)\left[-m + lD(t)\right]. \tag{3.10}$$

This last equation is now a decoupled linear system - recall $D(t)$ is diagonal - and can easily be solved. Going now back to $P(t)$, and with initial condition $P(0) = 1$, its solution is

$$P(t) = Q \exp\{-mt + l \int_0^{T_c} dt D(t)\} Q^{-1} = \exp\{-mt + l \int_0^{T_c} dt A(t)\}. \qquad (3.11)$$

Hence, the infection propagator reads

$$P = P(T_c) = \exp\{T_c[-m + l\langle A \rangle]\}. \qquad (3.12)$$

Matrix $\langle A \rangle$ is the average of the adjacency matrix over the whole period: $\langle A \rangle = \int_0^{T_c} dt A(t)/T_c$. The fact that infection propagator contains only the average adjacency matrix, which is invariant under time permutations, is due to the fact that no temporal correlations are present. In order to find the threshold, we must find the transmission rate value for which $\rho[P(T_c)] = 1$. From Equation 3.12 we immediately see that this leads to

$$l_c = \frac{m}{\rho[\langle A \rangle]}. \qquad (3.13)$$

We numerically check the accuracy of this solution for a simple network obeying the commutativity relation: a continuous version of model ER-T (see Section 2.5). We assume that, for each t, $A(t)$ is an instance of an Erdős-Rényi graph, with average degree $k(t)$, and average over all the possible instances. In this case, Equation 3.13 reduces to

$$l_c = \frac{m}{\bar{k}}, \qquad (3.14)$$

where $\bar{k} = \frac{1}{T_c} \int_0^{T_c} dt k(t)$, i.e., the average over time of the average degree. We choose a periodic $k(t)$:

$$k(t) = \frac{k_{max} - k_{min}}{2} \cos\left(\frac{2\pi}{T_r}t\right) + \frac{k_{max} + k_{min}}{2}, \qquad (3.15)$$

where k_{max}, k_{min} are the limit values of $k(t)$, and T_r is the period of $k(t)$. In Figure 3.2 we compare the l_c with the threshold obtained through discrete sampling (l_d), at different sampling intervals Δt. To be precise, the sampled network being discrete, it gives us a critical transmission probability λ_c. We then translate into a rate l_c, through the equivalence $\lambda_c = 1 - e^{-l_c \Delta t}$ [3]. When the sampling interval is comparable with the timescale of evolution T_r, large fluctuations of l_d compared to l_c, depending on the starting point of sampling. When instead sampling interval is significantly smaller than T_r, the two thresholds coincide, showing the accuracy of Equation 3.12.
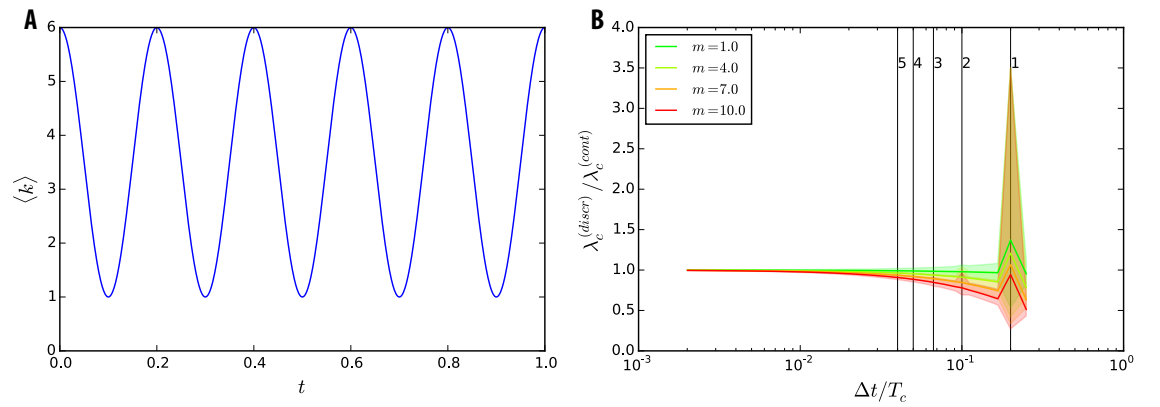
Figure 3.2.: **The epidemic threshold of model** ER-T: **discrete v continuous. A** shows the evolution of $k(t)$ for $T_r = 0.2, T_c = 1, k_{min} = 1, k_{max} = 6$. **B** has sampling interval on the x-axis, as a fraction of T_c, and on y-axis it has the ratio l_d/l_c, i.e., discrete threshold divided by continuous threshold. Four different values of recovery rate are explored, and for each of them we show the average across different starting points of sampling (solid line), and extreme values (colored areas). Vertical black lines mark multiples of T_r.

3.5. Latency

In this section we study the impact of latency period on the epidemic threshold. Specifically, we derive an analytical formula for computing the threshold of the SEIS model on a generic temporal network. For a description of the SEIS model, see Figure 1.1. When the network does not evolve in time, latency plays no role in shaping the epidemic threshold, and a SEIS model with parameters ϵ, μ will have the same threshold as a SIS model with parameter μ [146]. We investigate what happens for temporal networks. We have shown how crucial the interplay between network time scale and disease time scale (represented by $1/\mu$) is. We now wish to determine if the additional time scale $1/\epsilon$ induced by latency has an impact on the threshold and how we can compute it.

Before proceeding, we enumerate here the conceptual steps that will take us to the solution of the problem:

1. we write down the Markov chain equations of the SEIS model on a static network;

2. we design a new 2-layer static network, and couple it to a specifically designed SIS spreading dynamics;

3. we show that the two processes – SEIS on static network and SIS on 2-layer – have the same epidemic threshold;

4. we recover the known result that in the static case latency does not matter;

5. we extend this mapping to the temporal case, using the results of previous sections (Section 3.2 and [1]), and recover a formula for computing the epidemic threshold of the SEIS model on a generic temporal network.

We now develop in detail each step of the previous list.

1.

In analogy to what we did for the SIS model in Equation 1.20, we write the Markov chain equations of the SEIS model. We assume here a static unweighted, undirected network of N nodes, and of adjacency matrix A. We used to deal with a system of N equations, one for each node. Now we add $q_i(t)$, namely the probability node i is exposed, resulting

in $2N$ equations:

$$\begin{cases} p_i(t+1) = (1-\mu)p_i(t) + \epsilon q_i(t); \\ q_i(t+1) = (1-\epsilon)q_i(t) + [1 - p_i(t) - q_i(t)]\left\{1 - \prod_j [1 - \lambda A_{ij}p_j(t)]\right\}. \end{cases} \quad (3.16)$$
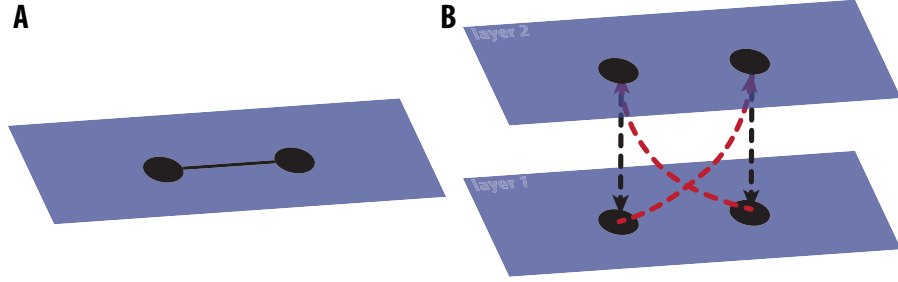
Figure 3.3.: **Schematic description of the 2-layer structure. A** represents the starting static network, containing here just two nodes and one link. **B** represents the 2-layer structure we define. Each node is present in both layers, and the copy in layer 2 points to the copy in layer 1 (black dashed arrows). The link in **A** is translated into two directed links from layer 1 to layer 2 (red dashed arrows).

2.

We now define a 2-layer static network, with each layer containing a copy of the N nodes. We do not allow for intra-layer links, so the resulting structure will be bipartite. Inter-layer links are defined following these two rules:

- each node in layer 2 is connected to its copy in layer 1 through a directed link (black dashed arrows in Figure 3.3**B**);

- if there is a link in the starting network ($A_{ij} = 1$), a directed link will go from i in layer 1 to j in layer 2, and another from j in layer 1 to i in layer 2 (red dashed arrows in Figure 3.3**B**).

These rules can be synthesized in terms of the adjacency tensor $\hat{A}_{\alpha\beta ij}$ of the 2-layer structure. $\hat{A}_{\alpha\beta ij} = 1$ means that node i in layer α is connected to node j in layer β, where $i, j = 1, \cdots, N$ (node indices) and $\alpha, \beta = 1, 2$ (layer indices). The components of

this tensor are

$$
\begin{cases}
\hat{A}_{11ij} = 0; \\
\hat{A}_{12ij} = A_{ij}; \\
\hat{A}_{21ij} = \delta_{ij}; \\
\hat{A}_{22ij} = 0.
\end{cases}
\tag{3.17}
$$

Switching to the supra-adjacency matrix formalism (Section 3.1) we write this tensor in terms of a $2N \times 2N$ adjacency matrix, divided into four blocks of dimension $N \times N$, with diagonal blocks showing intra-layer connectivity, and off-diagonal blocks inter-layer connectivity:

$$
\hat{A} = \left(\begin{array}{c|c} 0 & A \\ \hline 1 & 0 \end{array} \right)
\tag{3.18}
$$

We couple this new 2-layer structure to a SIS dynamic process, with heterogeneous transmission and recovery probabilities. In particular, we let each layer have its own recovery probability: μ_1, μ_2. In addition, we assume that transmission from layer 1 to layer 2 (occurring through red links in Figure 3.3**B**)) has a different probability than the opposite (black links): $\lambda_{12}, \lambda_{21}$. We then write the Markov chain equations of this process, which are a system of size $2N$, as the total number of nodes in $\hat{A}$. We conveniently call $p_i(t)$ the probability of node i being infectious in layer 1, and $q_i(t)$ of being infectious in layer 2.

$$
\begin{cases}
p_i(t+1) = (1-\mu_1)p_i(t) + \lambda_{21}\left[1 - p_i(t)\right]q_i(t); \\
q_i(t+1) = (1-\mu_2)q_i(t) + \left[1 - q_i(t)\right]\left\{1 - \prod_j \left[1 - \lambda_{12}A_{ij}p_j(t)\right]\right\}.
\end{cases}
\tag{3.19}
$$

We now compare Equation 3.16 (SEIS dynamic on A) and Equation 3.19 (SIS dynamics on $\hat{A}$), and set the parameters in the latter as follows: $\mu_1 = \mu$, $\mu_2 = \epsilon$, $\lambda_{12} = \lambda$, $\lambda_{21} = \epsilon$. We choose these value in order to make the two sets of equations as similar as possible. The result is

$$
\begin{cases}
p_i(t+1) = (1-\mu)p_i(t) + \epsilon\left[1 - p_i(t)\right]q_i(t); \\
q_i(t+1) = (1-\epsilon)q_i(t) + \left[1 - q_i(t)\right]\left\{1 - \prod_j \left[1 - \lambda A_{ij}p_j(t)\right]\right\}.
\end{cases}
\tag{3.20}
$$

3.

Equation 3.16 and Equation 3.20 look similar, but still have dissimilar terms. As a result, SEIS dynamics on A is not the same process as this SIS dynamics on $\hat{A}$. This

is not surprising, as the terms that make the equations different result from the fact that in the SEIS model a node can be either exposed or infectious; in the SIS model on 2-layer, on the contrary, a node can be infected in both layers at the same time. All dissimilar terms, however, are quadratic terms in p and q. Hence, when we linearize the two processes, i.e., Equation 3.16 and Equation 3.20, we find the same equations for both:

$$\begin{cases} p(t+1) = (1-\mu)p(t) + \epsilon q(t); \\ q(t+1) = \lambda A p(t) + (1-\epsilon)q(t) + \left\{ \mathcal{O}(||p||^2) + \mathcal{O}(||q||^2) + \mathcal{O}(||p||||q||) \right\}. \end{cases} \tag{3.21}$$

The two processes are different, but their dynamics close to the disease-free state is the same. As a result, since the threshold is computed exactly around that point, from the linearized equations, the two processes have the same epidemic threshold. We stress that this is not an approximate result: they have *exactly* the same threshold, as threshold is completely determined by the linearized behavior. The remarkable implication of this is that we can compute the threshold of the SEIS model by computing the threshold of a SIS model. We reduce model complexity, at the expense of topologic complexity, which however we can deal with.

4.

We can rewrite Equation 3.21 in matrix form. It suffice us to define a vector of dimension $2N$, by joining vectors p and q:

$$\begin{pmatrix} p(t+1) \\ \hline q(t+1) \end{pmatrix} = \begin{pmatrix} 1-\mu & \epsilon \\ \hline \lambda A & 1-\epsilon \end{pmatrix} \begin{pmatrix} p(t) \\ \hline q(t) \end{pmatrix}, \tag{3.22}$$

where the blocks are $N \times N$. As for the SIS model [155, 156], we get the threshold λ_c when the spectral radius of the matrix in Equation 3.22 is equal to one. This condition is equivalent to

$$\det \begin{pmatrix} -\mu & \epsilon \\ \hline \lambda A & -\epsilon \end{pmatrix} = 0. \ \Rightarrow \ \left(\frac{\lambda}{\mu} \right)_{critical} = \frac{1}{\rho[A]}. \tag{3.23}$$

As expected, this is the same threshold as in Equation 1.23.

5.

To get the threshold on a time-evolving network, we need to extend the mapping from A to $\hat{A}$ to the temporal case, and we achieve this using our previous results (Section 3.2

and [1]). The key is to be able to derive the new infection propagator $\hat{P}$. As usual, we start from a temporal network with snapshots $\{A_1, A_2, \cdots, A_T\}$. We then map each snapshot onto the 2-layer structure just described: $\{\hat{A}_1, \hat{A}_2, \cdots, \hat{A}_T\}$. At this point, this is a generic temporal network with $2N$ nodes, whose infection propagator we know (Equation 3.2). For the sake of clarity, here we rewrite Equation 3.2 and explicitly mark adjacency matrices, and their dimension as subscript: $\mathbb{I}_N$ is the $N \times N$ identity matrix. Moreover, we will use 0_N as the $N \times N$ matrix with all zero entries.

$$P = \prod_{h=1}^{T} [(1 - \mu)\mathbb{I}_N + \lambda A_h] \tag{3.24}$$

At this point we just need to explicitly account for heterogeneous transmission and recovery probabilities. Formally, we can do that through a few rules that translate the terms in the original infection propagator (Equation 3.24):

- $N \longrightarrow 2N$;

- λ [scalar] $\longrightarrow \left(\begin{array}{c|c} \lambda\mathbb{I}_N & 0_N \\ \hline 0_N & \epsilon\mathbb{I}_N \end{array} \right)$ [$2N \times 2N$ matrix];

- $(1 - \mu)$ [scalar] $\longrightarrow \left(\begin{array}{c|c} (1-\mu)\mathbb{I}_N & 0_N \\ \hline 0_N & (1-\epsilon)\mathbb{I}_N \end{array} \right)$ [$2N \times 2N$ matrix];

- A [$N \times N$ matrix] $\longrightarrow \left(\begin{array}{c|c} 0_N & A \\ \hline \mathbb{I}_N & 0_N \end{array} \right)$ [$2N \times 2N$ matrix]

By applying these transformations onto Equation 3.24, we get the new infection propagator

$$\hat{P} = \prod_{t=1}^{T} \left(\begin{array}{c|c} 1-\mu & \epsilon \\ \hline \lambda A_t & 1-\epsilon \end{array} \right) \tag{3.25}$$

Here we omit again identity and zero matrices, as their presence is now clear from the matrix algebra rules.

Once we know the infection propagator $\hat{P}$, we know the threshold of the SEIS model on temporal networks, using the results in [1], Section 3.2, Section 3.3:

$$\rho[\hat{P}(\hat{\lambda}_c, \epsilon, \mu)] = 1. \tag{3.26}$$

From now on, we indicate as λ_c the threshold of the SIS model, and as $\hat{\lambda}_c$ the threshold of the SEIS model, assuming same μ for both models. Through Equation 3.26, one

can compute the epidemic threshold of the SEIS model on a generic temporal network, which may generally be different from the SIS model, as the infection propagators are different. We devote the rest of this section to determining when λ_c and $\hat{\lambda}_c$ are indeed different. In order to highlight which are the topologic features that impact this difference, we will focus on network models, or equivalently classes of networks sharing some common properties, and for which infection propagators are averaged over the different realizations. We indicate such average with $\langle \cdot \rangle$. We know from Section 3.4 that, when no temporal correlations are present among edges ($\langle [A_t, A_s] \rangle \equiv 0 \; \forall t, s = 1, \cdots, T$), we can compute the threshold using the average adjacency matrix. The same happens with the SEIS model, but this time the adjacency matrix to consider is $\hat{A}$ (Equation 3.18). Whenever $\langle [\hat{A}_t, \hat{A}_s] \rangle \equiv 0$, then we can compute the threshold from $\langle \hat{A}_t \rangle$. But in this case, as we saw from Equation 3.23, latency factors out, and $\hat{\lambda} = \lambda$. Hence, $\langle [\hat{A}_t, \hat{A}_s] \rangle \equiv 0$ implies $\hat{\lambda} = \lambda$. We remark, however, that the condition $\langle [\hat{A}_t, \hat{A}_s] \rangle \equiv 0$ is much more stringent than the the absence of temporal correlations among edges ($\langle [A_t, A_s] \rangle \equiv 0$):

$$\left\langle [\hat{A}_t, \hat{A}_s] \right\rangle = \left(\begin{array}{c|c} \langle A_t \rangle - \langle A_s \rangle & 0 \\ \hline 0 & -(\langle A_t \rangle - \langle A_s \rangle) \end{array} \right). \tag{3.27}$$

Consequently, $\langle [\hat{A}_t, \hat{A}_s] \rangle \equiv 0$ if and only if $\langle A_t \rangle = A$, i.e., the average adjacency matrix does not change over time. This means that whenever every snapshot is sampled from the same ensemble, then $\hat{\lambda}_c = \lambda_c$. This is true only for models that do not exhibit temporal correlations, and whose parameters do not change over time, like ACT and ER (see Section 2.5). Any other network may exhibit $\hat{\lambda}_c = \lambda_c$. We can synthesize this condition as

$$\langle A_t \rangle \text{ constant} \implies \hat{\lambda}_c = \lambda_c. \tag{3.28}$$

The opposite implication is not true, as there may be particular situations in which $\hat{\lambda}_c$ is numerically the same as λ_c, even in presence of non-constant adjacency matrix.

3.6. Conclusion

In this chapter we have developed an analytic theory for computing the epidemic threshold on a temporal network, accounting for a completely general topology, including directed and weighted links. In addition, we have treated the different basic features the disease may exhibit, like the presence or absence of immunity, its duration, and a latency period. In Chapter 5 we will use this methodology to provide a quantitative

assessment of network vulnerability in different real scenarios, in terms of the risk that the introduction of a new pathogen will cause an epidemic outbreak.

4. Cattle trade movements

Diseases affecting farmed animals are extremely relevant in a Public Health perspective. In addition to compromising animal welfare, they represent a major cause of loss in economic revenue, in terms of loss of productivity, and extremely costly eradication plans [199, 200], due to resources directly allocated and livestock culling. In addition, they may directly impact human health, as most emerging human diseases have a zoonotic origin [201].

Due to its geographical position and integrated livestock market system, Europe has witnessed several emergencies due to outbreaks of livestock diseases. Among them we mention the notorious Foot-and-Mouth (FMD) outbreak in 2001 in UK [106, 199, 202]. Several other ailments represent a menace to European livestock market. Some threaten to reach the EU from neighboring regions, while some others are already endemic in certain EU countries, where they compromise productivity and from where they can spread further. We mention, among others, Bovine Viral Diarrhea (BVD) [203–205], Bovine Tuberculosis (BTB) [206–208], Brucellosis [209–211], affecting mostly cattle. Bluetongue [212], affecting mostly sheep, but also all ruminants. Classical Swine Fever [213], African Swine Fever [214], affecting swine. Despite their epidemiology being very diverse – Bluetongue and African Swine Fever, for instance, are vector-borne – their spread is driven, or at least facilitated, by livestock displacements among animal holdings, both within countries and across countries. As a result, studying how animals are displaced is a key step in devising new targeted prevention and containment strategies.

In every EU country, livestock displacements are now tracked[1] on a daily basis and, at least for cattle, at the animal level. Several studies have focused on specific countries and species [61–66]. Here we provide a comparative analysis of three EU countries: Hungary,

---

[1]European Parliament and European council (2000) 204: 1-10. Regulation (EC) No. 1760/2000 of 17 July 2000 establishing a system for the identification and registration of bovine animals and regarding labeling of beef and beef products and repealing Council Regulation (EC) No. 820/97 European Council. Off. J. Eur.Communities L.

Italy and Sweden. Swedish dataset has been used already in [63], while Hungarian dataset has never been analyzed before. Italian dataset was analyzed only for periods of just one year [64, 215–217]. Here we analyze a dataset of several consecutive years, which is also the basis of our work in Chapter 5 and [2]. This work is a fruit of a collaboration with the Hungarian National Food Chain Safety Office[2], the Italian Istituto Zooprofilattico "G. Caporale"[3], and Linköping University[4] in Sweden.

	holdings	movements/year	dataset
Italy	$1.7 \cdot 10^5$	$6.1 \cdot 10^6$	2006-2012
Hungary	$4.2 \cdot 10^4$	$3.8 \cdot 10^5$	2010-2013
Sweden	$2.5 \cdot 10^4$	$8.1 \cdot 10^5$	2008

Table 4.1.: **Average size of cattle markets.**

The interest in these countries lies in the fact that they pertain to different geographic regions, and supposedly have different market dynamics. Moreover, they are extremely heterogeneous in terms of number of holdings and trading volumes. Table 4.1 shows that Italy is the largest market, both in terms of number of premises, and displacements. Both Hungary and Sweden are almost one order of magnitude smaller. However, while Hungary has more premises than Sweden, its average traffic volume is smaller, meaning that Swedish premises are on average more active in exchanging animals than Hungarian ones.

4.1. Defining the network

Many studies represent cattle displacements as a network [62–66], and indeed such representation naturally arises from the data. We consider animal holdings as nodes of our network. Whenever holding i sends cattle to holding j following a commercial transaction, we represent this as a directed link from i to j, with weight equal to the number of animals sent. Given that these transactions are recorded on a daily basis, we have a discrete-time temporal network, whose snapshots represent displacements occurring

---

[2]www.nebih.gov.hu/en/

[3]www.izs.it/IZS/

[4]www.liu.se
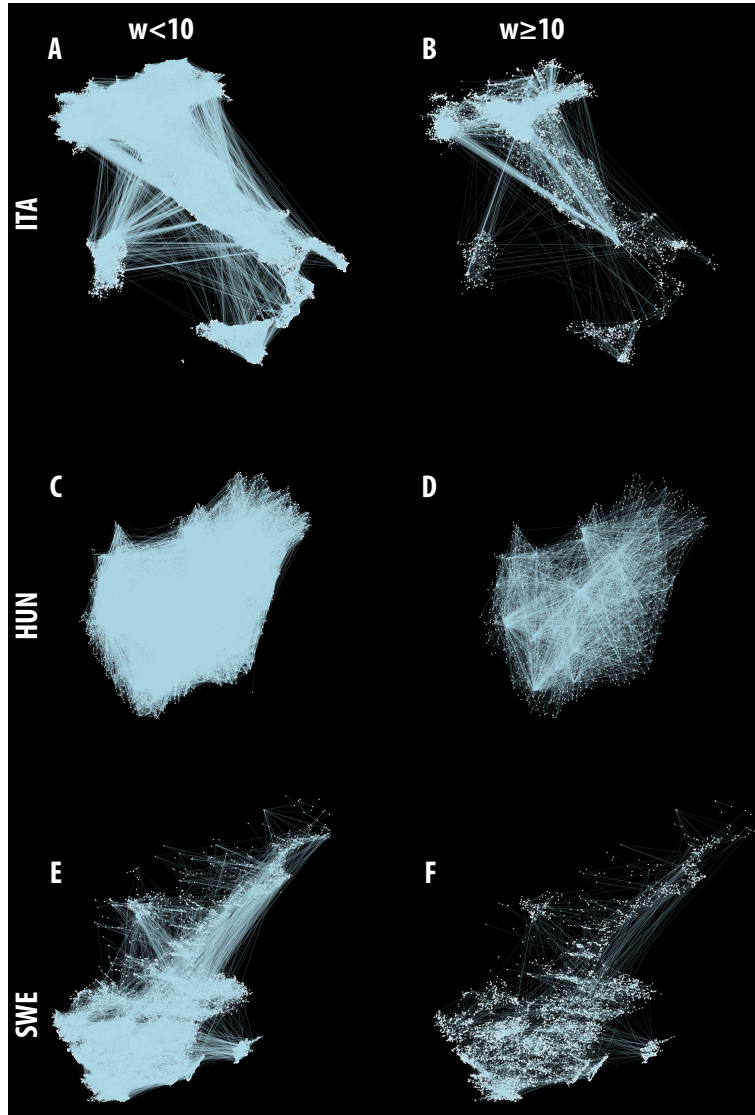
Figure 4.1.: **Yearly aggregated cattle networks.** Representation of cattle network aggregated over a year: 2008 for Italy (**A**,**B**) and Sweden (**E**,**F**), 2010 for Hungary (**C**,**D**). Nodes (animal holdings) are placed according to their geographic position. First column shows movements up to 10 animals (**A**,**C**,**E**); second column over 10 animals (**B**,**D**,**F**).

daily. Clearly, when we need a coarser view of the network, we can aggregate it in weekly, monthly and yearly networks. Following [64], we define months as periods of 28 days, so that each year contains 13 months. In addition to its topologic structure and weights, there are additional metadata that need to be included in some analyses. For instance, the network is embedded in physical space, as nodes represent real geographic locations. Figure 4.1 shows a yearly aggregated representation of the networks of the three countries, placing holdings according to their geographic coordinates.
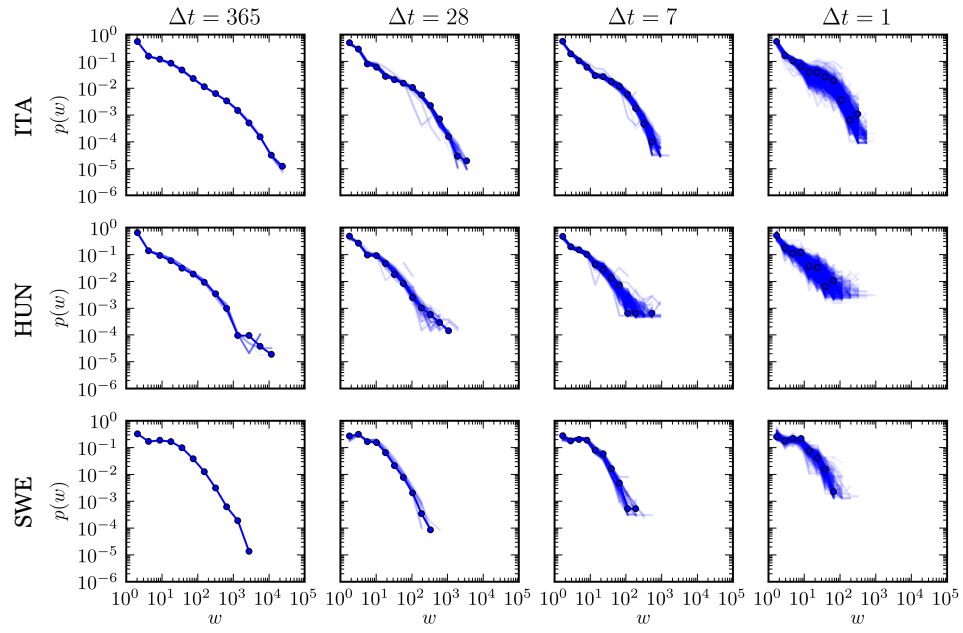
4.2. Static topologic properties
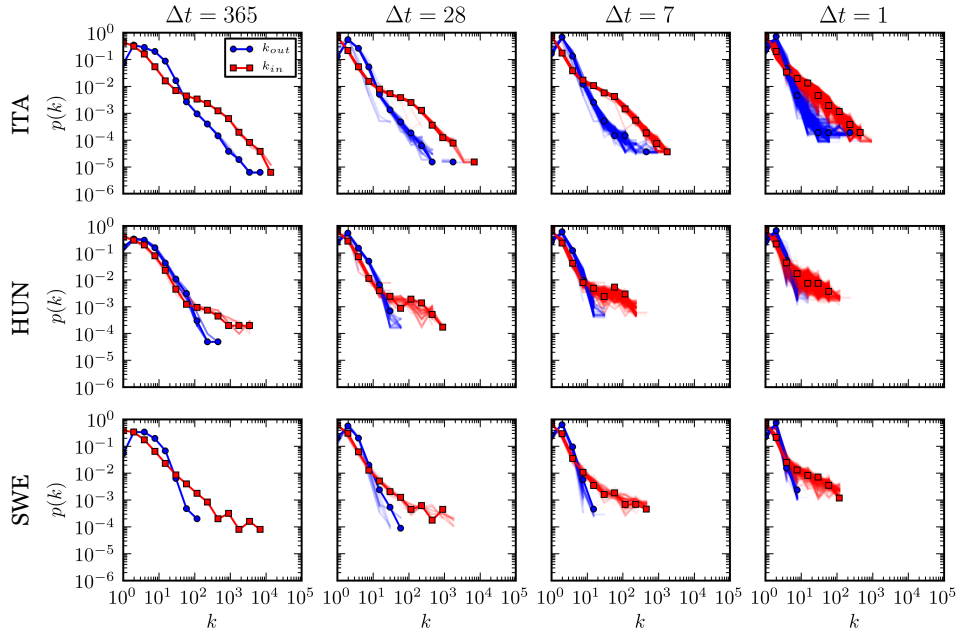
Figure 4.2.: **Degree distributions.** We plot degree distributions at different aggregation time scales: day ($\Delta t = 1$), week ($\Delta t = 7$), month ($\Delta t = 28$), year ($\Delta t = 365$). In-degree in in red, out-degree is in blue. For each aggregation intervals we plot in red squares or blue dots the distribution of one of such intervals, and in transparent red or blue all the others.

Centrality measures like node degree and strength are known to be proxies for infection
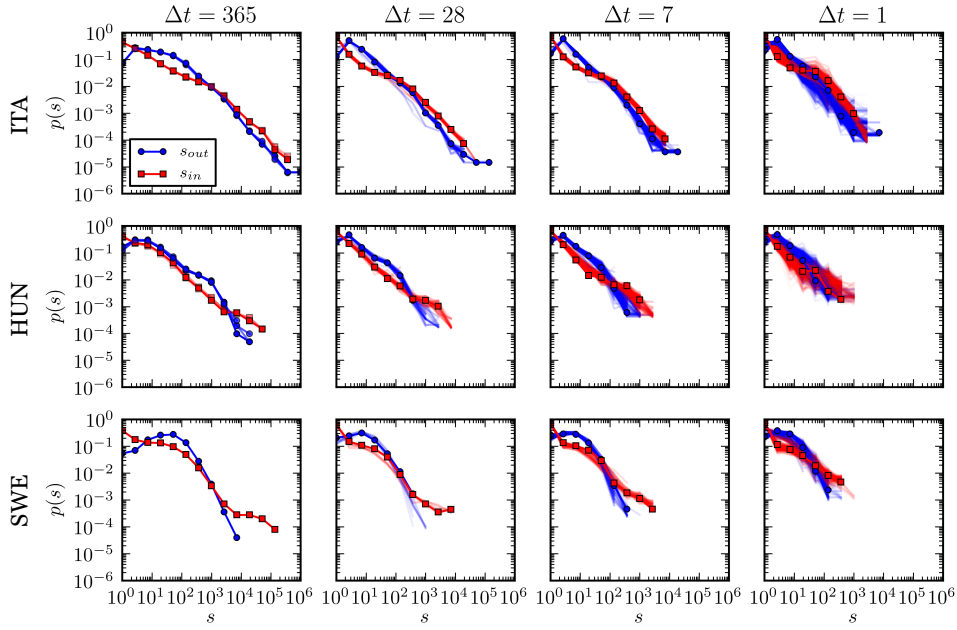
Figure 4.3.: **Strength distributions.** We plot strength distributions at different aggregation time scales: day ($\Delta t = 1$), week ($\Delta t = 7$), month ($\Delta t = 28$), year ($\Delta t = 365$). In-strength in in red, out-strength is in blue. For each aggregation intervals we plot in red squares or blue dots the distribution of one of such intervals, and in transparent red or blue all the others.

risk [30,76]. We analyze node degree and strength at different aggregation intervals: daily ($\Delta t = 1$), weekly ($\Delta t = 7$), monthly ($\Delta t = 28$) and yearly ($\Delta t = 365$). In Figure 4.2 we plot degree distribution, both for k_{in} (incoming links) and k_{out} (outcoming links). Three remarkable patterns emerge. Firstly, within the same country and aggregation intervals, distribution are extremely stable, as was already observed in [64] using one year of Italian movements. For instance, different yearly networks have the same degree distribution. Secondly, in each country, degree distributions appear heterogeneous across the different aggregation intervals. Unlike the activity driven model ACTIVITY, in which aggregated heterogeneity arises from homogeneous snapshots (see Section 2.5), here heterogeneous behavior is visible down to the highest temporal resolution, again as it was observed in [64]. This is particularly visible for in-degree, which is strongly heavier tailed than out-degree, meaning that hubs in terms of incoming traffic are more frequent, and larger, than hubs of outgoing traffic. Finally, despite system sizes varying greatly across countries, resulting in noisier distributions for Sweden and Hungary, we see qualitative properties seem to hold across countries as well. In all three indeed we observe heterogeneous distributions for in-degree. Out-degree is markedly heavy tailed in Italy, much less in the other two countries. This is dramatically visible for Sweden, where for $\Delta t = 365$ in-degree spans four order or magnitudes, while out-degree just two, and seems to peak around a particular value, with much smaller variance. Strength distributions exhibit the same properties as degree (see Figure 4.3), and this is confirmed by the behavior of the average strength at fixed degree (Fig 4.4). Indeed $\langle s(k)_{in} \rangle$ grows linearly with k, as expected in the case of no correlations between degree and strength, in all countries, at all aggregation intervals. On the other hand, $\langle s(k)_{out} \rangle$ seems to grow slightly superlinearly with degree, as a sign of a weak positive correlation: the more outgoing connections a node has, the more animals it sends along each of these contacts. This effect, however, is difficult to disentangle from statistical noise, and even if genuine, is extremely weak.

Finally, we consider link weight distributions (Figure 4.5). In Italy and Hungary they show a well defined heavy tailed behavior. In Sweden such heterogeneity is less marked. As degree and strength, they are stable across aggregation intervals.

4.3. Temporal patterns

In Chapter 2 we have described node and edge inter-activation time as an important property of temporal networks. Here we compute its distribution in the three countries,
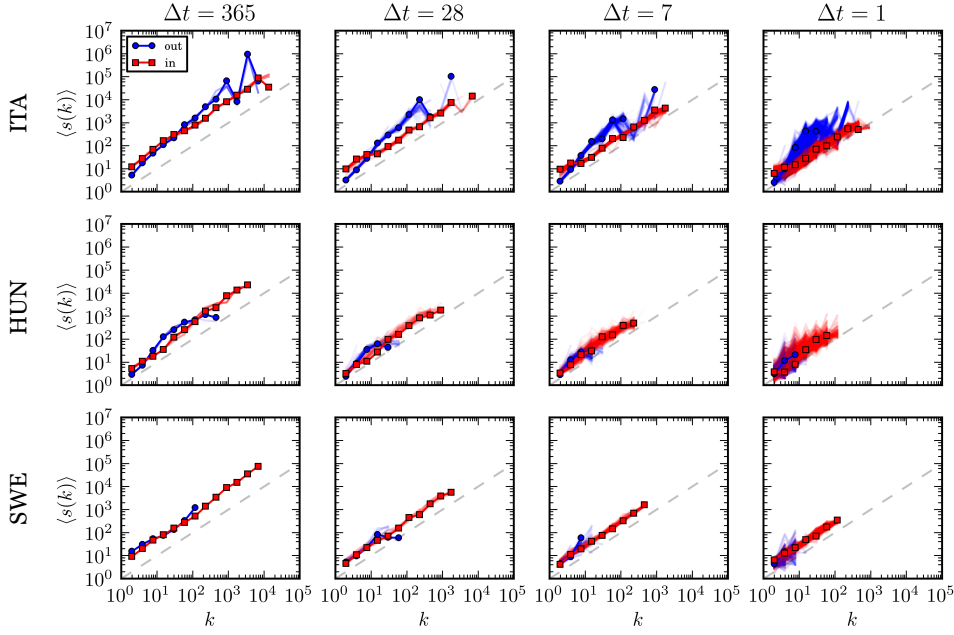
Figure 4.4.: **Strength v degree.** For each aggregation interval and each country we plot the average strength of nodes with a given degree. We examine different aggregation time scales: day ($\Delta t = 1$), week ($\Delta t = 7$), month ($\Delta t = 28$), year ($\Delta t = 365$). In-degree and in-strength are in red, out-degree and out-strength are in blue. For each aggregation intervals we plot in red squares or blue dots the distribution of one of such intervals, and in transparent red or blue all the others. Gray dashed lines indicate the linear relation $\langle s(k) \rangle = k$.

Figure 4.5.: **Link weight distributions.** We plot weight distributions at different aggregation time scales: day ($\Delta t = 1$), week ($\Delta t = 7$), month ($\Delta t = 28$), year ($\Delta t = 365$). For each aggregation intervals we plot in red blue dots the distribution of one of such intervals, and in transparent blue all the others.

both at daily (Figure 4.6) and weekly (Figure 4.7) aggregation interval. A clear weekly patterns emerges in both edge and node activation. Spikes in Figure 4.6 occur every 7 days, and indicate that a link (node) is more likely to re-activate after a time that is a multiple of the week. This is visible in all countries, and more marked for edges than for nodes. The fact that weekly aggregation does not show any recurrent pattern (Figure 4.7), but just heterogeneous distributions, shows that no time scale in network is as strong as the week, as far as inter-activation time is concerned.
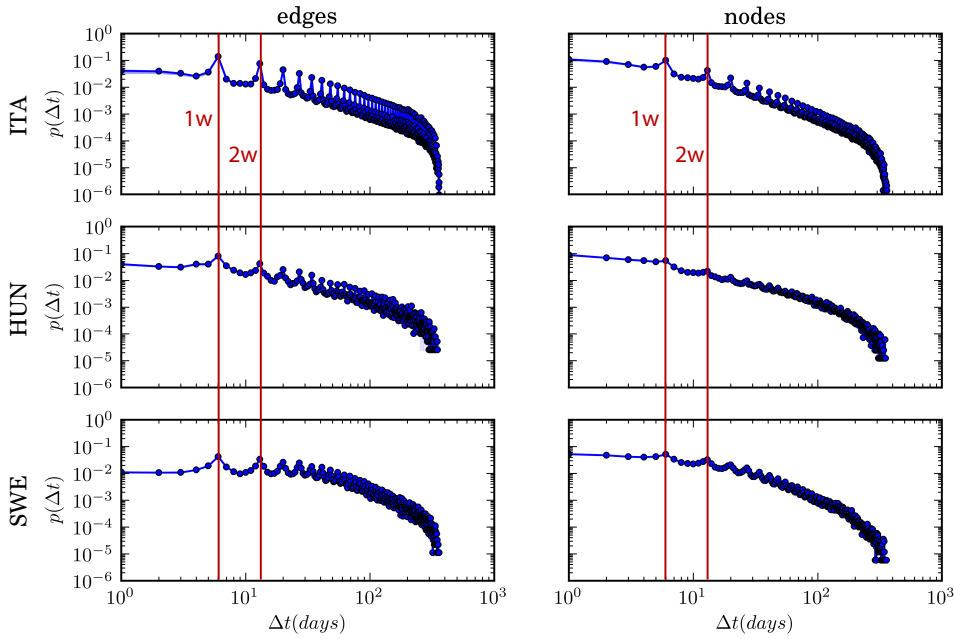
Figure 4.6.: **Inter-activation time distributions, daily network.** Inter-activation time distribution of edges (first column) and nodes (second column), on the daily aggregated network. Red vertical lines indicate an inter-activation time of one and two weeks.

We also compute the distribution of activation time, i.e. the time a node or link stays active without interruption. In Figure 4.8 we plot activity time distributions for weekly aggregation. They have a clean heavy tailed shape, again showing the deep heterogeneous behavior of these networks. Unlike inter-activation time, activation time distribution at daily aggregation (not shown) exhibits no recurrent patterns.
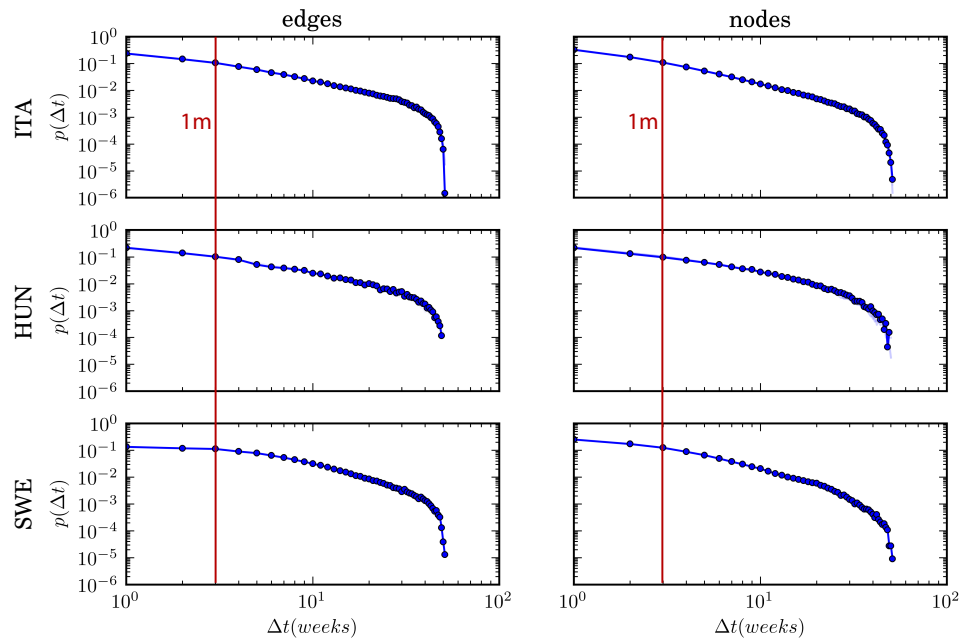
Figure 4.7.: **Inter-activation time distributions, weekly network.** Inter-activation time distribution of edges (first column) and nodes (second column), on the weekly aggregated network. Red vertical line indicates an inter-activation time of one month.
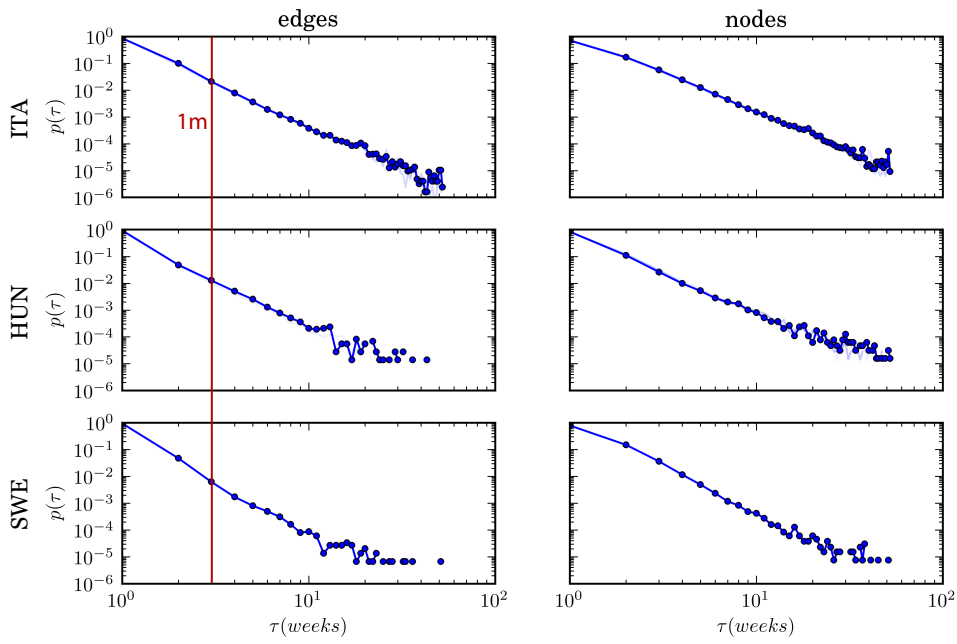
Figure 4.8.: **Activation time distributions, weekly network.** Activation time distribution of edges (first column) and nodes (second column), on the weekly aggregated network, i.e., the number of consecutive weeks a edge (node) stays active. Red vertical line indicates an inter-activation time of one month.

As the week emerges as a fundamental time scale of evolution in all three countries, we wish to examine it more in detail, by looking at the characteristic activity of each day of the week. Figure 4.9 shows the number of displaced animals by day of the week. We see that, in all countries, weekdays are more active than weekends, as one could expect. Such difference is however extremely marked in Italy, where Mondays, on average, witness $3.5 \cdot 10^4$ displacements, compared to Sundays with less than $5 \cdot 10^3$. In Sweden, and particularly in Hungary, such difference is less pronounced. In addition Italy is the only country with a noticeable difference between Saturday and Sunday (the latter being the least active) and between Monday (the most active day) and the other weekdays.
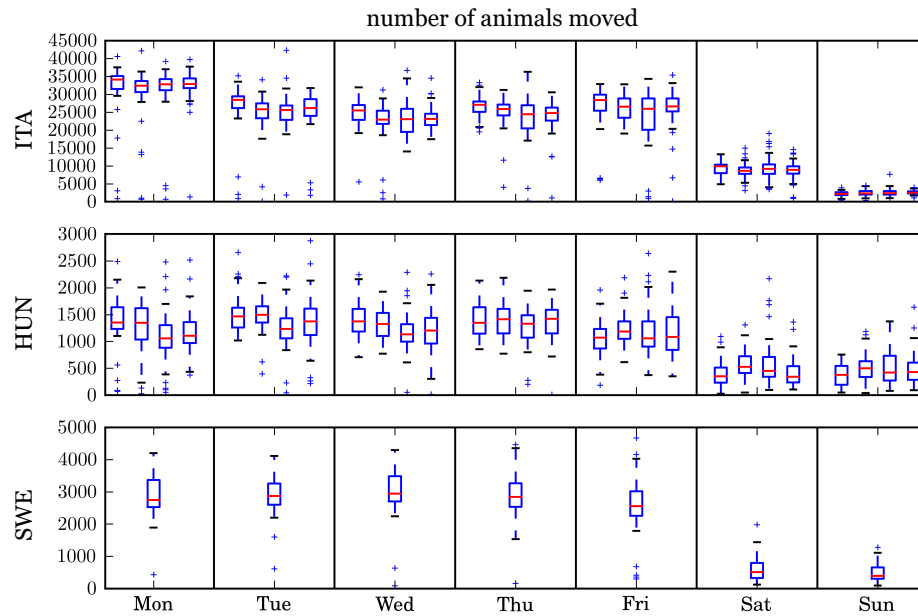
Figure 4.9.: **Activity by day of the week.** We plot traffic (i.e., number of displaced animals) for each day of the week. Boxplots represent median, quartiles and outliers, and are computed each over one year. In analogy with Hungary, for Italy we show four years.

We then analyze activity timelines, i.e., the variation of number of cattle displacements in time. (Figure 4.10). Sweden has a strong seasonal pattern, with peaks of activity

around May and October. They are both preceded by a trough around March and July. Italy's seasonal variation is visible too, albeit less marked. Specifically, year 2008 seems to behave differently than the others, with a more pronounced seasonal behavior. The activity peak in fall occurs later than in Sweden. Both Italy and Sweden have a trough corresponding to Christmas and the beginning of January. Italy has a pronounced trough for the week of August 15, too. Hungary's timeline shows no seasonal behavior. We find our three countries to have diverse behaviors as far as seasonality is concerned, and their different climatic conditions probably contribute to that effect. In analogy with Hungary, for Italy we show just four years.
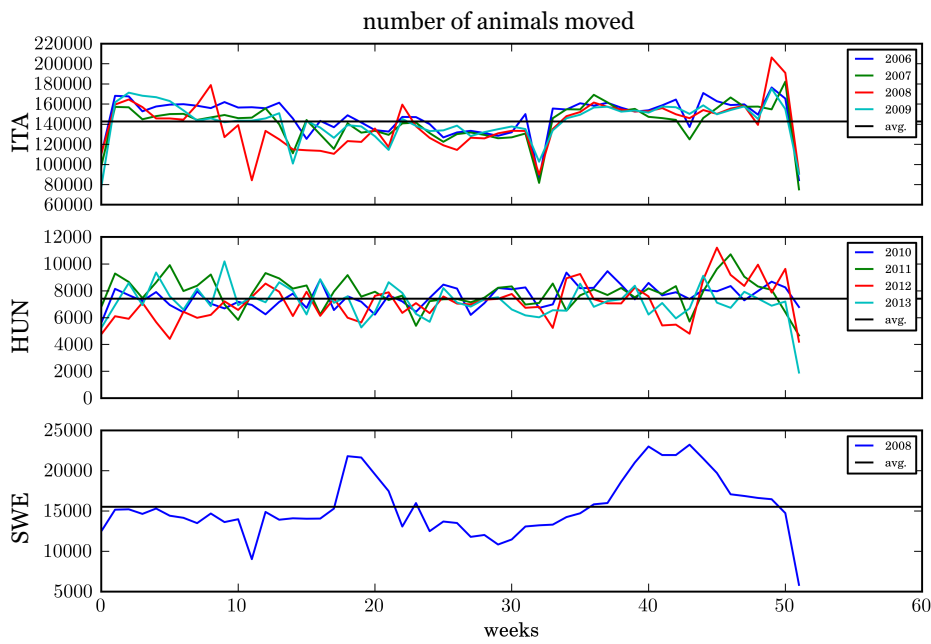
Figure 4.10.: **Activity timelines.** Weekly activity timelines (traffic per week) in different countries and different years, and their respective average (black line). Highest peaks represent 60% relative variation to the mean for Italy, 50% for Hungary, and 65% for Sweden. In analogy with Hungary, for Italy we show four years.

So far we have uncovered the presence of two time scales, the first being the week (Figure 4.6 and 4.9), the second being the year (Figure 4.10). We now wish to rigorously recover them inside the activity timeline. In order to do that we need to deal

with the daily timeline, which is however too noisy for a visual analysis like we have done for the weekly timeline. Hence, we use discrete Fourier transform (DFT) to sistematically uncover all the relevant time scales (see Appendix A.3 for a description of DFT). In Figure 4.11 we plot the intensity of each time scale (blue crosses). Then we apply RESHUFFLE null model (see Section 2.6) onto each network several times, and for each instance we compute the DFT. Purple lines correspond to the 95% confidence interval of the intensity of that time scale in the null models. Consequently, real values standing above the upper purple line represent a statistically significant time scale, with respect to the uncorrelated null models. As expected, all countries show marked peaks corresponding to the weekly time scale, which is definitely the strongest. Italy has a statistically significant yearly time scale, while Hungary does not, confirming the absence of seasonality. We are not able to compute the yearly harmonics for Sweden, since we have just one year of dataset, but still we find a significant 6 month timescale, which we were not able to spot by just looking at the activity timeline (Figure 4.10). Finally all networks have significant peaks at 1/2 of a week and 1/3 of a week (except Hungary), and these are again higher harmonics of the weekly pattern which we could not pinpoint without spectral analysis.

Time scales, however, inform us about the overall macroscopic trend of the network, but do not delve into its microscopic dynamics. For example, if we observe that same months of different years have roughly the same number of links, we may wonder if those links that activate are actually the same from one year to the other, i.e., if the network retains memory of its structure. To uncover this fact, we compute the microscopic memory of network snapshots. In [2] we did it for yearly snapshots of the Italian dataset, and in [64] the same analysis is performed for monthly snapshot inside one year. Here we expand to monthly snapshots of several years and all three countries. As a measure of memory we choose *overlap*, i.e., the fraction of common edges between two snapshots σ, τ:

$$\text{overlap}(\sigma, \tau) = \frac{|\mathcal{E}(\sigma) \cap \mathcal{E}(\tau)|}{|\mathcal{E}(\sigma) \cup \mathcal{E}(\tau)|}, \tag{4.1}$$

where $\mathcal{E}(\sigma)$ is the set of links in snapshot σ. Clearly, when overlap= 0, the two snapshots share no links, while overlap= 1 means the two snapshot have exactly the same network structure. The higher is the overlap, the more τ has memory of what links were active in σ (assuming $\tau > \sigma$). Figure 4.12 shows the value of overlap among monthly snapshots. We remark that memory is never high: snapshot never share more than a fraction 0.22
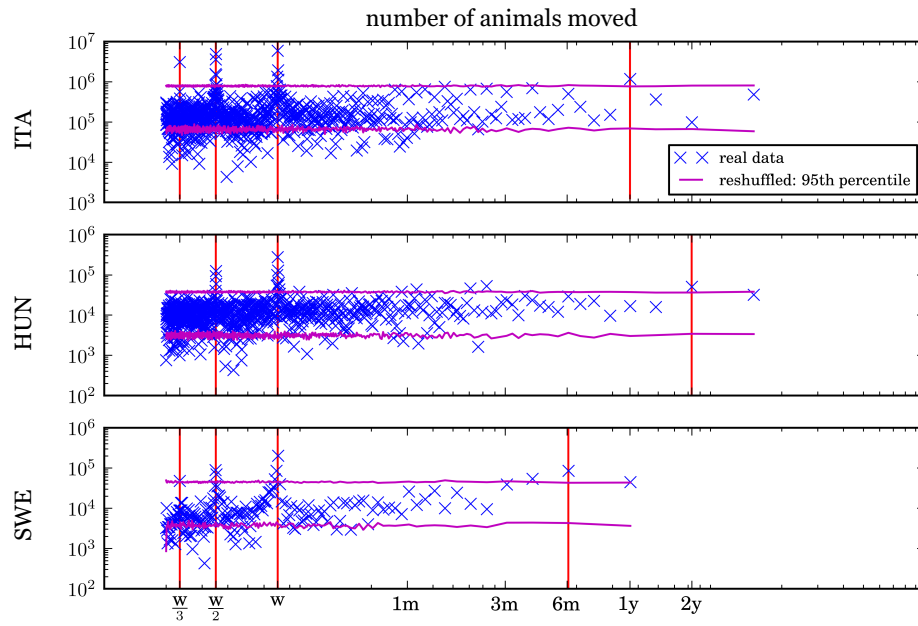
Figure 4.11.: **Spectral analysis of activity timelines.** Harmonic intensities of daily activity timelines of the full three datasets (blue crosses). Purple lines are the $95th$ percentile of the harmonic intensity of the RESHUFFLE null model, as defined in Section 2.6.

of their structure. In addition, memory is significantly higher in Italy and Sweden, than Hungary. Given that for Sweden we have only one year, it is difficult to pinpoint any pattern, and we just observes that the first four months tend to be more similar among themselves then average. Italy and Hungary, for which we can go further than one year, show remarkably different behaviors. In Italy we observe two distinct patterns. Firstly, months of the same year tend to be more similar among themselves than average, forming clusters of similarity. We can see this in Figure 4.12A as diagonal blocks of purple. Secondly, and more interestingly, we see that the yearly pattern is visible in memory, too, in form of purple lines parallel to the diagonal in Figure 4.12A. This means that monthly snapshot distant exactly one, two or three years are more similar than average. In Hungary we observe the former pattern, and in particular we see that the first year (2010) is very different from all the others. We do not observe the latter. As a result, in Italy we see a yearly periodicity both in macroscopic activation patterns, and microscopic similarity. Finally, we see neither in Sweden.

4.4. Geographic patterns

In the last step of our analysis, we make use of the fact that animal holdings have a specific location in geographic space, and so nodes of our networks are geo-referenced. We wish to find out wether within the examined countries there are regions that concentrate most of the activity of cattle market, or wether activity is homogeneously distributed. The representation of the aggregated network of Figure 4.1 already hints to the fact that such heterogeneities might be present. We divide the national territories in small squares, and for each of them we plot the outgoing (Figure 4.13**A**,**C**,**E**) and incoming (Figure 4.13**B**,**D**,**F**) traffic. We uncover that while Italy and Hungary have active traffic to and from most part of their national territory, in Sweden farming is concentrated in the Southern and coastal areas. On the other hand, heterogeneities in traffic volume among active areas are extremely marked in Italy and Hungary, than Sweden. In Italy traffic is concentrated in the Northern part, and particularly in the *Pianura Padana* around Cuneo, Lombardy, Emilia-Romagna, and Veneto. Other localized hotspot are visible, especially as incoming traffic, around Rome, Naples and Syracuse. In Hungary Figure 4.13**B**,**D** show several localized hotspots on a rather uniform background. In Sweden the picture is different, as active areas tend to have uniform traffic, both incoming and outgoing.
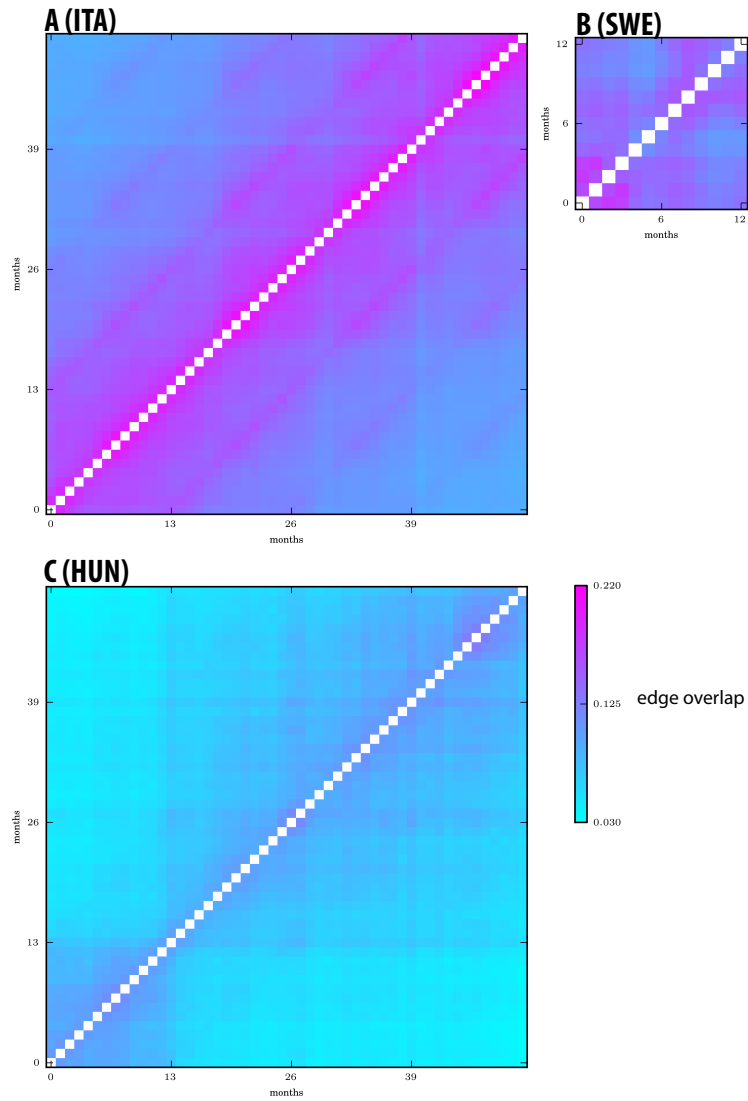
Figure 4.12.: **Memory.** Overlap of monthly snapshots (Equation 4.1) in Italy (**A**), Sweden (**B**), Hungary (**C**). In analogy with Hungary, for Italy we show four years.
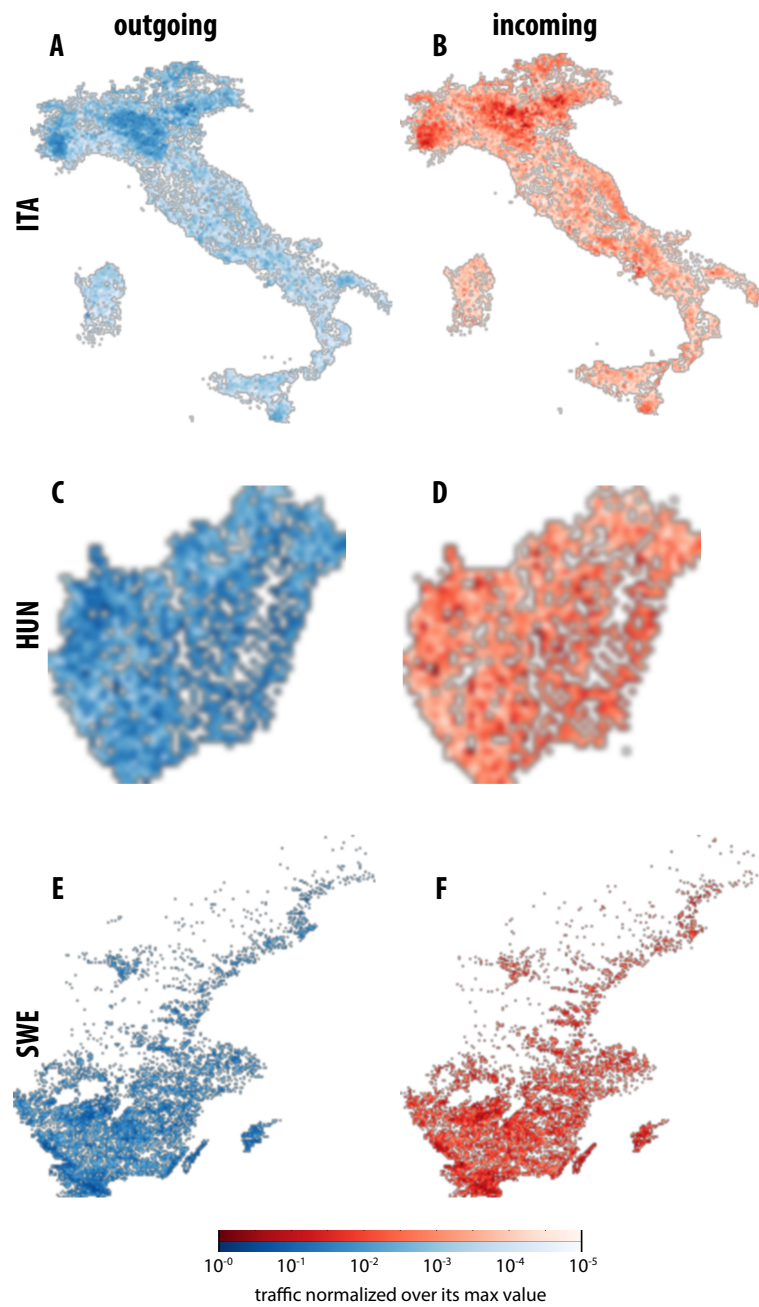
Figure 4.13.: **Spatial patterns of cattle traffic.** Outgoing (blue scale) and incoming (red scale) traffic volume across the territory of Italy (**A**,**B**), Hungary (**C**,**D**), Sweden (**E**,**F**). In order to highlight heterogeneities, traffic value is normalized over its highest value in the country.

4.5. Conclusion

In conclusion, cattle trade networks exhibit an extremely rich and diverse phenomenology in terms of aggregated properties, temporal structures, and geographic patterns. Several further analyses could be performed, including more metadata, like farm production type, or animal species. We have shown that at the aggregated level networks in different countries qualitatively share the same features, mostly in terms of heavy tailed distribution of centrality measures. We also have found that week is the dominating time scale in all three countries, while seasonal and yearly behaviors in both activation and memory tend to be country-specific. In addition, we have found that spatial patterns are country-specific, too, as they respond to differences in climate and geography.

In Chapter 5 we will focus on the Italian network, and assess its vulnerability of to pathogen introduction, using our computation of the epidemic threshold (Chapter 3). We will also develop a node-centered risk prediction methodology, able to compute the risk of a specific holding being hit by the disease in the future, by relying only on past network data (Section 5.3).

5. Targeted risk prediction

In Chapter 3 we developed a theoretical framework for computing the epidemic threshold on temporal networks, for several disease models. We now wish to apply it to real case studies, as a practical tool to assess the vulnerability of a system to pathogen introduction. We choose to consider the Italian network of cattle displacements (Chapter 4), and the network of human sexual interactions (Section 2.4). Both these systems are extremely relevant from the epidemiological point of view. Cattle network represents a substrate for the diffusion of many animal diseases that can potentially harm both animal health and economy, as we have explained at the beginning of Chapter 4. Sexual contact network allows us to study spreading dynamics of several sexually transmitted infections.

Using the epidemic threshold, we study how the vulnerability of the cattle trade network changes across space and time (Section 5.1). We then apply the theoretical findings of Section 3.5 to the sexual contact network, and assess the impact of latency period on the epidemic threshold, exploring several parameter configurations, and in particular the ones that correspond to syphilis.

Computing epidemic risk through epidemic threshold has however one limitation, due to the fact that we often have contact data concerning a limited amount of time, and do not now how representative they are of network future evolution. In the last part of this chapter, which consists in article [2], we overcome this by putting forward a methodology to predict epidemic risk that relies only on the knowledge of past contact data. Moreover, while threshold computation is a global measure of vulnerability, here we compute a targeted epidemic risk, consisting in the probability a specific node will be hit during a future outbreak. While the disease model we consider is the very simple SI, we explain in [2] that it is a reasonable approximation for the initial stage of the outbreak, and test its validity for different parameter values (Supporting Material of [2]). The methodology we present is entirely data-driven, and must be therefore tested case by case. We show, however, that it is extremely general, and apply it to both cattle trade movements and

sexual contacts. Finally, we develop a synthetic network model to uncover which are the relevant structural features behind the performance of our methodology.

5.1. Vulnerability of Italian cattle trade network

In this section we study the global and regional vulnerability of the Italian cattle trade network, described in Section 4. We adopt the commonly used assumption [63, 202, 205, 217] of considering animal holdings as epidemiological units, which can transmit the disease by sending out infected cattle. We assume an average infectious period $(1/\mu)$ of 28 days. Thanks to what we found in Section 3.3, the resulting epidemic threshold holds for SIS, SIR, and $SIRS$ models. Even though these models are a simplification of realistic disease progressions, they nonetheless cover a large variety of disease types, so that the resulting vulnerability assessment can be adapted to different scenarios. The SIS framework, for instance, has already been adopted for modeling Bovine Viral Diarrhea (BVD) [205], while the SIR model has been adapted to Foot and Mouth [202].

For every year, from 2006 to 2012, we compute the epidemic threshold of this network, taken at its highest temporal resolution (1 day). Result is shown in Figure 5.1. In addition, we consider regional subnetworks and compute their threshold. Given a region, a regional subnetwork is defined as the network containing all animal holdings (nodes) present in that region, and all the links among them. In this way, measuring the epidemic threshold on a region means assessing the local vulnerability, discounting long range movements. Figure 5.2 shows the epidemic thresholds on the coarser administrative units in Italy – *regioni* –, corresponding to a European *NUTS2* division[1]. In Figure 5.3 instead we subdivide Italy according to *NUTS1* territorial units, which consist in five macro-areas with no administrative meaning. We observe a variation of the national threshold among the seven years considered. Specifically, values are clustered around three values, as follows: $2006, 2007; 2008, 2010, 2011; 2009; 2012$. Regional subnetworks in the Northern part of the country exhibit temporal fluctuations, too. Additionally, we observe a clear geographical pattern. Subnetworks in the North are much more vulnerable (lower threshold) than in the Center and South, where many regions are below threshold for every possible transmissibility (white). We explain these sharp geographic differences with the different structure and nature of livestock market. In the *Pianura*

---

[1] *NUTS – Nomenclature des unités territoriales statistiques*, is a EU subdivision standard defined by Eurostat.

Padana, a plane crossing Northern Italy from West to East, Southern of the Alps, farming is highly intensive and industrialized, with large active holdings. This feature emerged clearly in the analysis carried out in Chapter 4, and is visible in Figures 4.1,4.13. In the Central and Southern parts of Italy, apart some localized spots, farming is of smaller scale, with small holdings that exchange animals less often and in smaller quantities. This result indicates that only the Northern part of the country is able to sustain an endemic outbreak, provided the disease has no other way of spreading but animal movements, as it is the case, for instance for BVD [203–205]. This does not mean farms in the South are not at risk of being infected, but that the persistence of the virus in the country is due to transport activity in the North. In addition, the South may still be very important for the importation of exotic pathogens, due to its location and favorable environmental conditions.
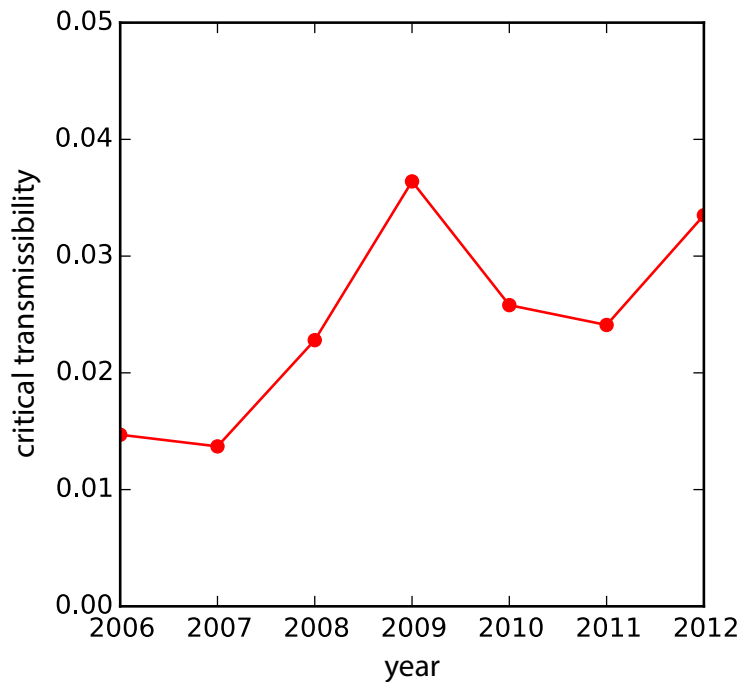
Figure 5.1.: **Yearly epidemic thresholds (2006-2012) of the Italian cattle trade network.** Average recovery rate is 28 days.

The temporal variations of the epidemic threshold are more difficult to explain, and will need further investigation. However, they are likely due to disrupting events that
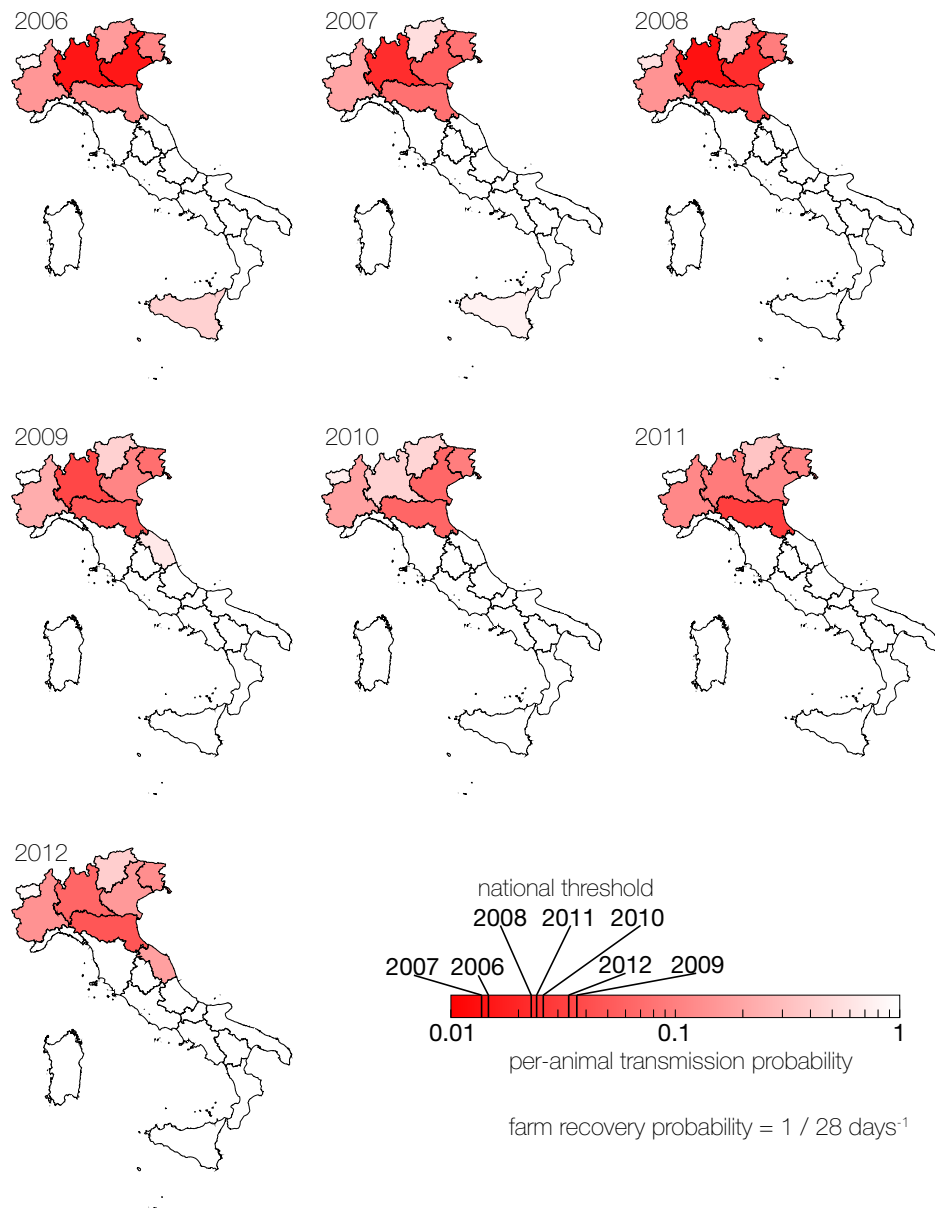
Figure 5.2.: **Yearly epidemic thresholds (2006-2012) of subnetworks corresponding to Italian regions.** Average recovery rate is 28 days. Regions are color coded according to their threshold. Thresholds of the full network are marked on the color bar.
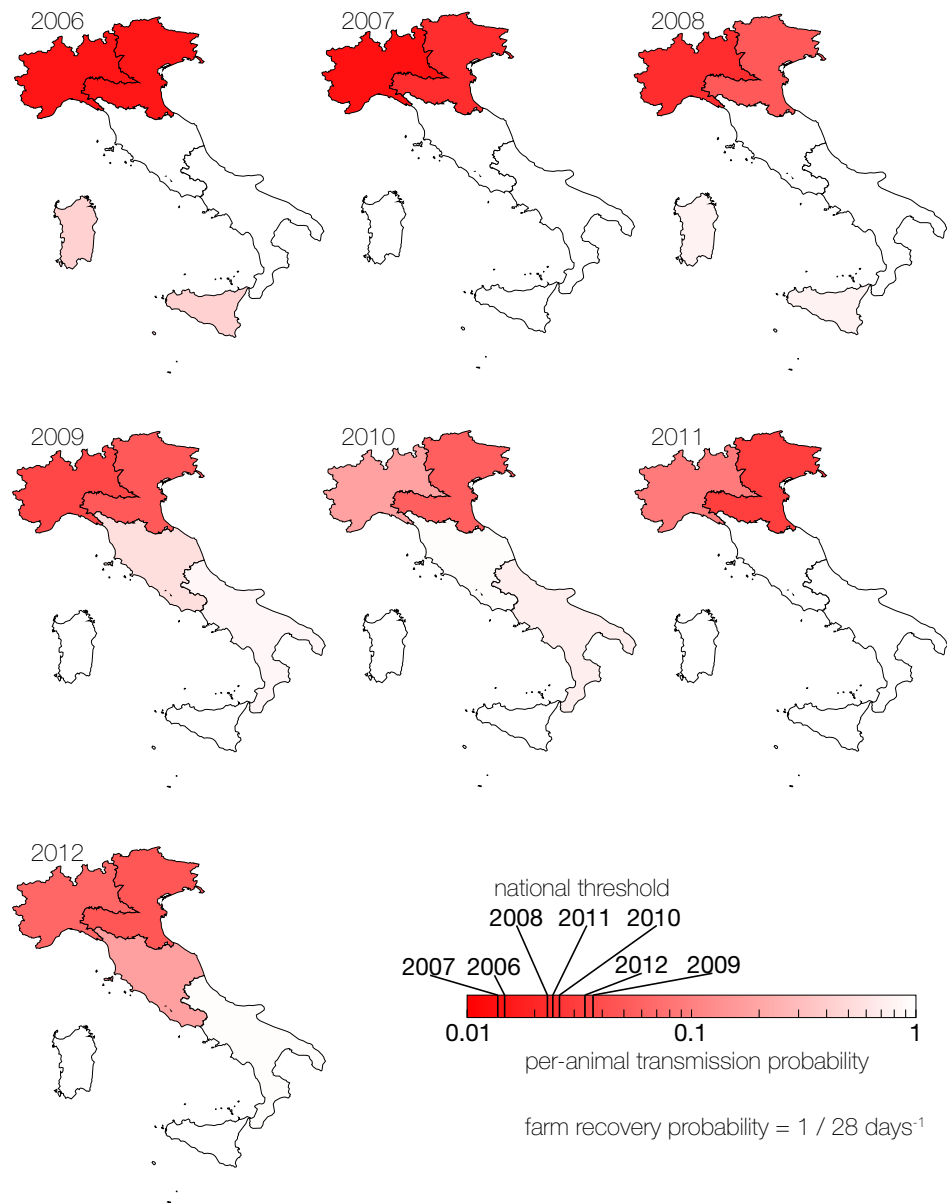
Figure 5.3.: **Yearly epidemic thresholds (2006-2012) of the subnetworks corresponding to NUTS1 geographic subdivision.** Average recovery rate is 28 days. NUTS1 are color coded according to their threshold. Thresholds of the full network are marked on the color bar.

qualitative change network topology, like outbreaks, or adoption of new laws. To determine this, we can discount the influence of trivial variations in the overall activity of the network. In Figure 5.4 we use average yearly node strength as a proxy for network activity, and correlate it with the value of the national epidemic threshold. We observe that relative variations in average strength are small ($< 10\%$), and cannot explain the observed variability in threshold. This means that network topology qualitatively changes from one year to another.
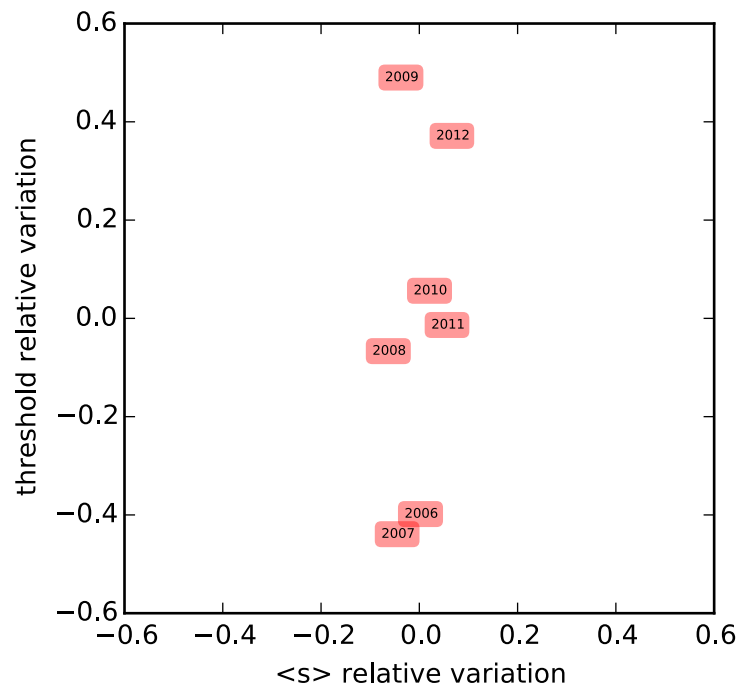
Figure 5.4.: **Yearly cattle Italian networks: relative variations in average node strength v relative variations in threshold.** The average across years of the average node strength is computed, and relative variations around it are plotted in the x-axis. Relative variations around the average yearly threshold are potted in the y-axis.

5.2. Sexual contact network: how latency influences vulnerability

In Section 3.5 we have analytically shown that the interplay between latency period and network evolution impacts the value of the epidemic threshold. In addition, we provided a formula for computing the threshold when latency is accounted for. Here we wish to apply these theoretical findings to the spread of STDs on a sexual contact network (see [60] and Section 2.4). For the same value of recovery probability μ, we compute the threshold of the SIS (λ_c) and SEIS ($\hat{\lambda}_c$) models. We wish to assess for which parameter values latency has a significant impact on the epidemic threshold, and whether that impact makes the network more or less vulnerable to the pathogen, with respect to the model without latency.

We explore different average infectious periods ($1/\mu$), and different average latency periods $1/\epsilon$, and compute the relative difference between the SEIS and SIS thresholds, in terms of

$$\Xi = \frac{\hat{\lambda}_c - \lambda_c}{\lambda_c}. \tag{5.1}$$

Ξ quantifies the impact of latency on the epidemic threshold. $\Xi \approx 0$ means that its impact is negligible. Instead, when $\Xi > 0$, latency increases the threshold, making the network less vulnerable to the disease. In the opposite regime ($\Xi < 0$) latency lowers the threshold, making the network more vulnerable. Results are shown in Figure 5.5. Figure 5.5**A** plots Ξ as a function of latency period, at fixed values of infectious period. Figure 5.5**B** comprehensively explores parameter space, with latency and infectious periods ranging from few days to almost one year. We observe a rich phenomenology, with Ξ ranging from around -5% to 40%. Latency in turn represents a risk factor ($\Xi < 0$, red in Figure 5.5**B**), plays no role ($\Xi \approx 0$, white in Figure 5.5**B**), or is a protection factor ($\Xi > 0$, blue in Figure 5.5**B**), depending on parameter values.

Finally, we focus on a specific sexually transmitted disease, syphilis, and in particular on its primary stage. This has a latency period ranging from around 7 to 90 days [218] (black box in Figure 5.5**B**). Within this interval, we find all three scenarios. For infectious periods around 15 days, we find latency has no impact on the epidemic threshold. For shorter periods, latency tends to lower the threshold with respect to the pure SIS model, especially for short infectious periods. Finally, for latency periods larger than 15 days, latency causes the threshold to be higher. Natural progression of syphilis is long and
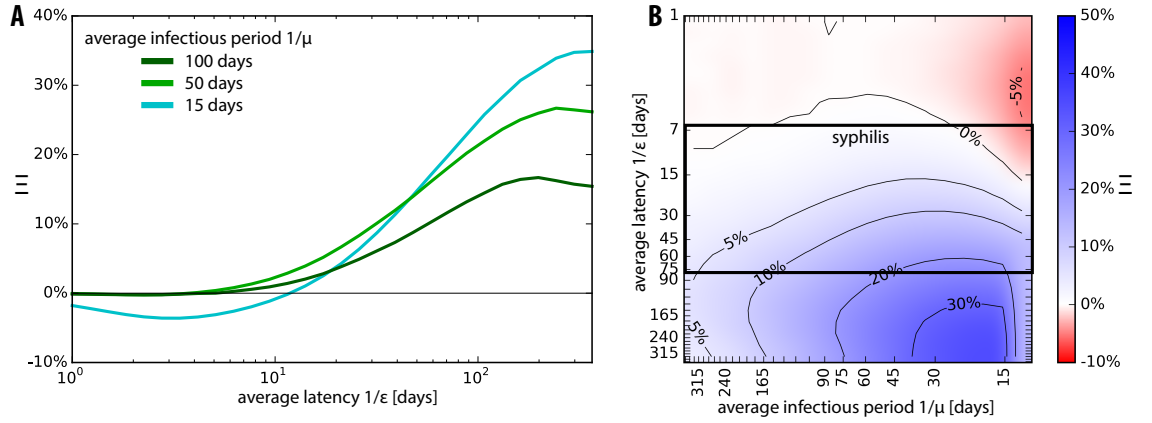
Figure 5.5.: **Sexual contact network:** $\hat{\lambda}_c$ **v** λ_c. In **A** we fix three different average infectious periods, explore the value of the average latency period (x-axis), and plot the value of Ξ (defined in Eq. 5.1) in the y-axis. In **B** we explore the full parameter space of both infectious and latency periods. Color code correspond to Ξ. Red indicates $\Xi < 0$, blue $\Xi > 0$. Black square corresponds to the latency period range for syphilis.

complex, and involves different stages, beyond the modeling capabilities of a SEIS model. However, given that *treponema pallidum* – bacteria responsible for syphilis – is sensitive to penicillin [218], we interpret the average infectious period as the average time it takes to being cured, once primary stage is symptomatic.

5.3. Article: Predicting epidemic risk from past temporal contact data

BEGINNING

Valdano E, Poletto C, Giovannini A, Palma D, Savini L, Colizza V

Predicting Epidemic Risk from Past Temporal Contact Data

Predicting Epidemic Risk from Past Temporal Contact Data

Eugenio Valdano [1,2], **Chiara Poletto** [1,2], **Armando Giovannini** [3], **Diana Palma** [3]
Lara Savini [3], **Vittoria Colizza** [1,2,4]*

1 INSERM, UMR-S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013 56 bd Vincent Auriol—CS 81393-75646 Paris Cedex 13, France, **2** Sorbonne Universités, UPMC Univ Paris 06, UMR-S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013 56 bd Vincent Auriol—CS 81393 - 75646 Paris Cedex 13, France, **3** Istituto Zooprofilattico Sperimentale Abruzzo-Molise G. Caporale Campo Boario, 64100 Teramo, Italy, **4** ISI Foundation Via Alassio 11/c, 10126 Torino, Italy

* vittoria.colizza@inserm.fr

Abstract

Understanding how epidemics spread in a system is a crucial step to prevent and control outbreaks, with broad implications on the system's functioning, health, and associated costs. This can be achieved by identifying the elements at higher risk of infection and implementing targeted surveillance and control measures. One important ingredient to consider is the pattern of disease-transmission contacts among the elements, however lack of data or delays in providing updated records may hinder its use, especially for time-varying patterns. Here we explore to what extent it is possible to use past temporal data of a system's pattern of contacts to predict the risk of infection of its elements during an emerging outbreak, in absence of updated data. We focus on two real-world temporal systems; a livestock displacements trade network among animal holdings, and a network of sexual encounters in high-end prostitution. We define the node's loyalty as a local measure of its tendency to maintain contacts with the same elements over time, and uncover important non-trivial correlations with the node's epidemic risk. We show that a risk assessment analysis incorporating this knowledge and based on past structural and temporal pattern properties provides accurate predictions for both systems. Its generalizability is tested by introducing a theoretical model for generating synthetic temporal networks. High accuracy of our predictions is recovered across different settings, while the amount of possible predictions is system-specific. The proposed method can provide crucial information for the setup of targeted intervention strategies.

Author Summary

Following the emergence of a transmissible disease epidemic, interventions and resources need to be prioritized to efficiently control its spread. While the knowledge of the pattern of disease-transmission contacts among hosts would be ideal for this task, the continuously changing nature of such pattern makes its use less practical in real public health emergencies (or otherwise highly resource-demanding when possible). We show that in

such situations critical knowledge to assess the real-time risk of infection can be extracted from past temporal contact data. An index expressing the conservation of contacts over time is proposed as an effective tool to prioritize interventions, and its efficiency is tested considering real data on livestock movements and on human sexual encounters.

Introduction

Being able to promptly identify who, in a system, is at risk of infection during an outbreak is key to the efficient control of the epidemic. The explicit pattern of potential disease-transmission contacts has been extensively used to this purpose in the framework of theoretical studies of epidemic processes, uncovering the role of the pattern's properties in the disease propagation and epidemic outcomes [1, 2, 3, 4, 5, 6, 7, 8]. These studies are generally based on the assumption that the entire pattern of contacts can be mapped out or that its main properties are known. Although such knowledge would be a critical requirement to conduct risk assessment analyses in real-time, which need to be based on the updated and accurate description of the contacts relevant to the outbreak under study [9], it can hardly be obtained in reality. Given the lack of such data, analyses generally refer to the most recent available knowledge of contact data, implicitly assuming a non-evolving pattern.

The recent availability of time-resolved data characterizing connectivity patterns in various contexts [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22] has inevitably weakened the non-evolving assumption, bringing new challenges to the assessment of nodes' epidemic risk. Traditional centrality measures used to identify vulnerable elements or influential spreaders for epidemics circulating on static networks [1, 2, 4, 23, 24, 25, 26, 27, 28, 29, 30] are unable to provide meaningful information for their control, as these quantities strongly fluctuate in time once computed on the evolving networks [19, 31]. An element of the system may thus act as *superspreader* in a past configuration of the contact network, having the ability to potentially infect a disproportionally larger amount of secondary contacts than other elements [32], and then assume a more peripheral role in the current pattern of contact or even become isolated from the rest of the system [19]. If the rules driving the change of these patterns over time are not known, what information can be extracted from past contact data to infer the risk of infection for an epidemic unfolding on the current (unknown) pattern?

Few studies have so far tried to answer this question by exploiting temporal information to control an epidemic through targeted immunization. They are based on the extension to temporal networks [33, 34] of the so-called acquaintance immunization protocol [4] introduced in the framework of static networks that prescribes to vaccinate a random contact of a randomly chosen element of the system. In the case of contacts relevant for the spread of sexually transmitted infections, Lee et al. showed that the most efficient protocol consists in sampling elements at random and vaccinating their latest contacts [33]. The strategy is based on local information gathered from the observation and analysis of past temporal data, and it outperforms static-network protocols. Similar results are obtained for the study of face-to-face contact networks relevant for the transmission of acute respiratory infections in a confined setting, showing in addition that a finite amount of past network data is in fact needed to devise efficient immunization protocols [34].

The aim of these studies is to provide general protocols of immunization over all possible epidemiological conditions of the disease (or class of diseases) under study. For this reason, protocols are tested through numerical simulations and results are averaged over starting seeds and times to compare their performance. Previous work has however shown that epidemic

outcomes may strongly depend on the temporal and geographical initial seed of the epidemic [35], under conditions of large dynamical variability of the network and absence of stable structural backbones [19]. Our aim is therefore to focus on a specific epidemiological condition relative to a given emerging outbreak in the population, resembling a realistic situation of public health emergency. We focus on the outbreak initial phase prior to interventions when facing the difficulty that some infected elements in the population are not yet observed. The objective is to assess the risk of infection of nodes to inform targeted surveillance, quarantine and immunization programs, assuming the lack of knowledge of the explicit contact pattern on which the outbreak is unfolding. Knowledge is instead gathered from the analysis of the full topological and temporal pattern of past data (similarly to previous works [33, 34]), coupled, in addition, with epidemic spreading simulations performed on such data under the same epidemiological conditions of the outbreak under study. More specifically, we propose an egocentric view of the system and assess whether and to what extent the node's tendency of repeating already established contacts is correlated with its probability of being reached by the infection. Findings obtained on past available contact data are then used to predict the infection risk in the current unknown epidemic situation. We apply this risk assessment analysis to two large-scale empirical datasets of temporal contact networks—cattle displacements between premises in Italy [19, 36], and sexual contacts in high-end prostitution [16]—and evaluate its performance through epidemic spreading simulations. We also introduce a model to generate synthetic time-varying networks retaining the basic mechanisms observed in the empirical networks considered, in order to explain the results obtained by the proposed risk assessment strategy within a general theoretical framework.

Results and Discussion

The cattle trade network is extracted from the complete dataset reporting on time-resolved bovine displacements among animal holdings in Italy [19, 36] for the period 2006–2010, and it represents the time-varying contact pattern among the 215,264 premises composing the system. The sexual contact network represents the connectivity pattern of sexual encounters extracted from a Web-based Brazilian community where sex buyers provide time-stamped rating and comments on their experiences with escorts [16].

The five-years data of the livestock trade network show that stationary properties at the global level co-exist with an active non-trivial local dynamics. The probability distributions of several quantities measured on the different yearly networks are considerably stable over time, as e.g. shown by the in-degree distribution reported in Fig. 1A, where the in-degree of a farm measures the number of premises selling cattle to that farm. These features, however, result from highly fluctuating underlying patterns of contacts, never preserving more than 50% of the links from one yearly configuration to another (Fig. 1C), notwithstanding the seasonal annual pattern due to repeating cycles of livestock activities [37, 38] (see S1 Text). Similar findings are also obtained for the sexual contact network (Fig. 1B-D), where the lack of an intrinsic cycle of activity characterizing the system leads to smaller values of the overlap between different configurations (< 10%). In this case we consider semi-annual configurations, an arbitrary choice that allows us to extract six network configurations in a timeframe exhibiting an approximately stationary average temporal profile of the system, after discarding an initial transient time period from the data [16]. Different time-aggregating windows are also considered (see the Materials and Methods section and S1 Text for additional details).
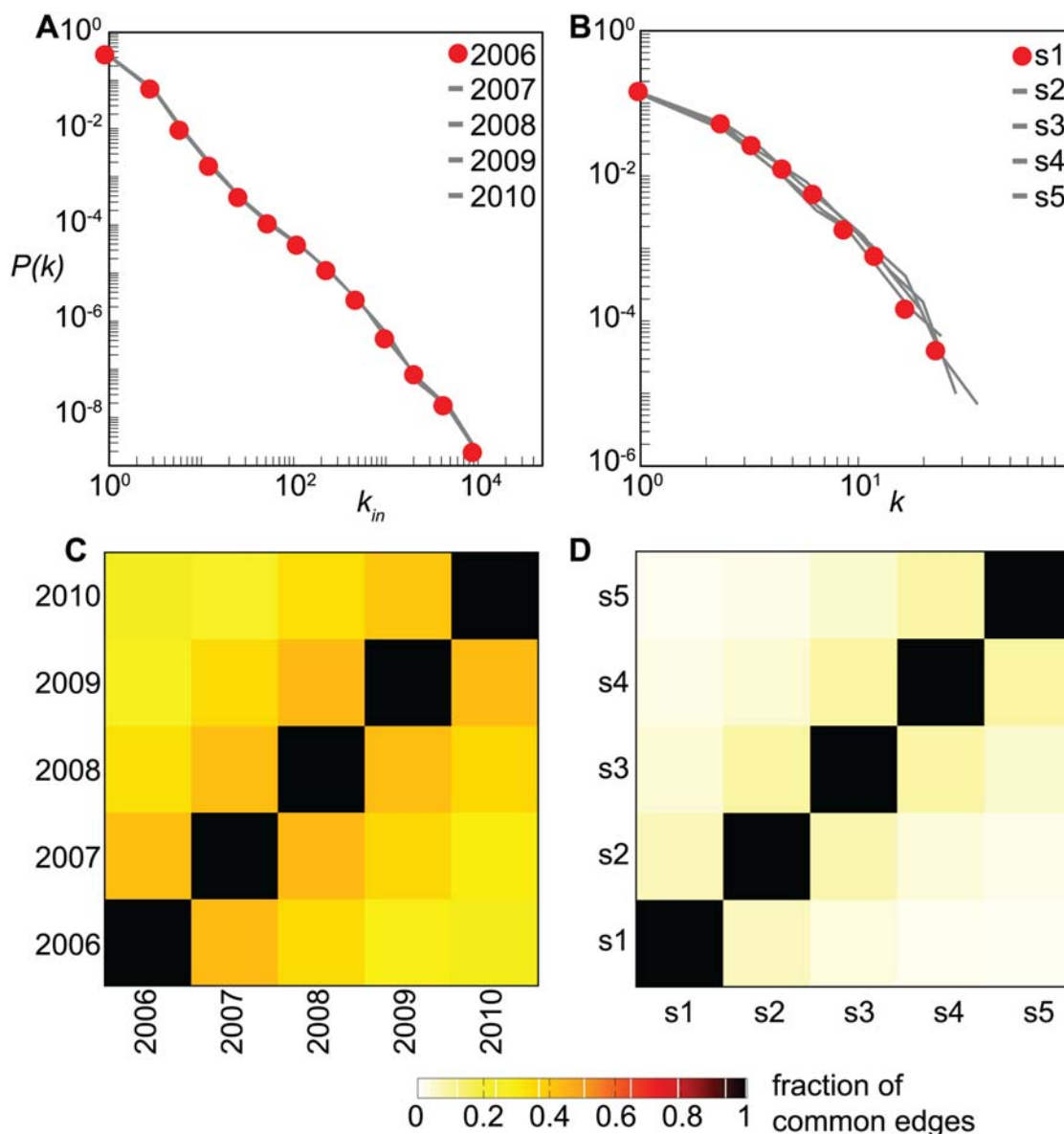
Fig 1. Structural and temporal properties of the cattle trade network and of the sexual contact network. (*A*), (*B*): premises in-degree distributions in the cattle trade network and sex customers degree distribution in the sexual contact network, respectively. Distributions for different configurations of the networks are superimposed in both cases. (*C*), (*D*): fraction of common edges contained in two configurations of the network, for the cattle trade network and the sexual contact network, respectively. In (*B*), (*D*) s stands for semester, the aggregation interval of each configuration.

doi:10.1371/journal.pcbi.1004152.g001

Loyalty

The observed values of the overlap of the time-resolved contact networks in terms of the number of links preserved are a measure of the degree of memory contained in the system. This is the outcome of the temporal activity of the elements of the system that reshape up to 50% or 90% of the contacts of the network (in the cattle trade case and in the sexual contact case, respectively), through nodes' appearance and disappearance, and neighborhood restructuring. By framing the problem in an egocentric perspective, we can explore the behavior of each single node of the system in terms of its tendency to remain active in the system and re-establish

connections with the same partners vs. the possibility to change partners or make no contacts. We quantitatively characterize this tendency by introducing the *loyalty* θ, a quantity that measures the fraction of preserved neighbors of a node for a pair of two consecutive network configurations in time, $c-1$ and c. If we define $\mathcal{V}_i^{c-1}$ as the set of neighbors of node i in configuration $c-1$, then $\theta_i^{c-1,c}$ is given by the Jaccard index between $\mathcal{V}_i^{c-1}$ and $\mathcal{V}_i^c$:

$$\theta_i^{c-1,c} = \frac{\left|\mathcal{V}_i^{c-1} \cap \mathcal{V}_i^c\right|}{\left|\mathcal{V}_i^{c-1} \cup \mathcal{V}_i^c\right|}. \tag{1}$$

Loyalty takes values in the interval [0, 1], with $\theta = 0$ indicating that no neighbors are retained, and $\theta = 1$ that exactly the same set of neighbors is preserved ($\mathcal{V}_i^{c-1} = \mathcal{V}_i^c$). It is defined for discrete time windows (c, $c+1$) and in general it depends on the aggregation interval chosen to build network configurations.

In case the network is directed, as for example the cattle trade network, θ can be equivalently computed on the set $\mathcal{V}_{in,i}^c$ of incoming contacts or on the set of neighbors of outgoing connections, $\mathcal{V}_{out,i}^c$, depending on the system-specific interpretation of the direction and on the interest in one phenomenon or the opposite. This measure originally finds its inspiration in the study of livestock trade networks, where a directed connection from holding A to holding B indicates that B purchased a livestock batch from A, which was then displaced along the link direction $A \rightarrow B$. If we compute θ on the incoming contacts of the cattle trade network, we thus quantify the propensity of each farmer to repeat business deals with the same partners when they purchase their cattle. This concept is at the basis of many loyalty or fidelity programs that propose explicit marketing efforts to incentivize the reinforcement of loyal buying behavior between a purchasing client and a selling company [39], and corresponds to a principle of exclusivity in selecting economic and social exchange partners [40, 41]. Analogously, in the case of the sexual contact network we consider the point of view of sex buyers. Formally, our methodology can be carried out with the opposite point of view, by considering out-degrees with loyalties being computed on out-neighbors. Our choice is arbitrary and inspired by the trade mechanism underlying the network evolution.

Other definitions of similarity to measure the loyal behavior of a node are also possible. In S1 Text we compare and discuss alternative choices. For the sake of clarity all symbols and variables used in the article are reported in Table 1. Finally, other mechanisms different from fidelity strategies may be at play that result in the observed behavior of a given node. In absence of additional knowledge on the behavior underlying the network evolution, we focus on the loyalty θ to explore whether it can be used as a possible indicator for infection risk, as illustrated in the following subsection.

The distributions of loyalty values, though of different shapes across the two datasets, display no considerable variation moving along consecutive pairs of configurations of each dataset (Fig. 2A-B and S1 Text), once again indicating the overall global stability of system's properties in time and confirming the results observed for the degree. A diverse range of behaviors in establishing new connections vs. repeating existing ones is observed, similarly to the stable or exploratory strategies found in human communication [42]. Two pronounced peaks are observed for $\theta = 0$ and $\theta = 1$, both dominated by low degree nodes for which few loyalty values are allowed, given the definition of Eq. (1) (see S1 Text for the dependence of θ on nodes' degree and its analytical understanding). The exact preservation of the neighborhood structure ($\theta = 1$) is more probable in the cattle trade network than in the sexual contact network ($P(\theta = 1)$ being one order of magnitude larger), in agreement with the findings of a higher system-wide memory reported in Fig. 1. Moreover, the cattle trade network exhibits the presence of high loyalty values (in the range $\theta \in [0.7, 0.9]$), differently from the sexual contact network where P

Table 1. List of variables and their description.

Notation	Description
c	index for network configurations
θ or $\theta_i^{c-1,c}$	loyalty of node i between configurations $c-1$, c
$\mathcal{V}_{in,i}^c$, $\mathcal{V}_{out,i}^c$	set of in(out)-neighbors of i in config c
L, D	loyalty classes (loyal, disloyal)
ε	loyalty threshold
s	epidemic seed
τ	duration of the outbreak early stage
$\mathcal{I}_s^c$	set of infected nodes for outbreak starting from s in config c
$\pi_D^{c-1,c}(s)$, $\pi_L^{c-1,c}(s)$	infection potentials for class D (L) computed for seed s between configs $c-1$, c
k	degree (in-degree for the cattle trade network)
$T_{DD}^c(k)$, $T_{DL}^c(k)$, $T_{LD}^c(k)$, $T_{LL}^c(k)$	transition probability from one loyalty class to another
ρ_i^c	epidemic risk for node i in config c
$\mathcal{I}_{s,h}^c$, $\mathcal{I}_{s,l}^c$	set of infected nodes with high(low) epidemic risk
P_h, P_l	probability of a high(low) risk node to be infected
$v = P_h/P_l$	risk ratio between P_h, P_l, measure of accuracy
$\omega_s^{c-1,c}$	predictive power (fraction of infected nodes for which it is possible to compute the epidemic risk)
b, d	node probability of becoming active or inactive
p_α	node probability of keeping an in-neighbor
α	number of kept in-neighbors
β_{in}	number of new in-neighbors
β_{out}	number of new out-neighbors
γ, δ	exponents of the distributions of β_{in}, β_{out}

doi:10.1371/journal.pcbi.1004152.t001

(θ) is always equal to zero in that range except for one pair of consecutive configurations giving a positive probability for $\theta = 0.8$. Farmers in the cattle trade network thus display a more loyal behavior in purchasing cattle batches from other farmers with respect to how sex buyers establish their sexual encounters in the analyzed sexual contact dataset.

For the sake of simplification, we divide the set of nodes composing each system into the subset of *loyal nodes* having θ greater than a given threshold ϵ, and the subset of *disloyal nodes* if instead $\theta < \epsilon$. We call hereafter these classes as *loyalty statuses L* and *D*, respectively, and we will later discuss the role of the chosen value for ϵ.

Epidemic simulations and risk of infection

Both networks under study represent substrates offering potential opportunities for a pathogen to diffuse in the corresponding populations. Sexually transmitted infections spread among the population of individuals through sexual contacts [43, 44], whereas livestock infectious diseases (e.g. Foot-and-mouth disease [45], Bluetongue virus [46], or BVD [47]) can be transmitted from farm to farm mediated by the movements of infected animals (and vectors, where relevant), potentially leading to a rapid propagation of the disease on large geographical scales.

As a model for disease-transmission on the network of contacts we consider a discrete-time Susceptible-Infectious compartmental approach [48]. No additional details characterizing the course of infection are considered here (e.g. recovery dynamics), as we focus on a simplified theoretical picture of the main mechanisms of pathogen diffusion and their interplay with the
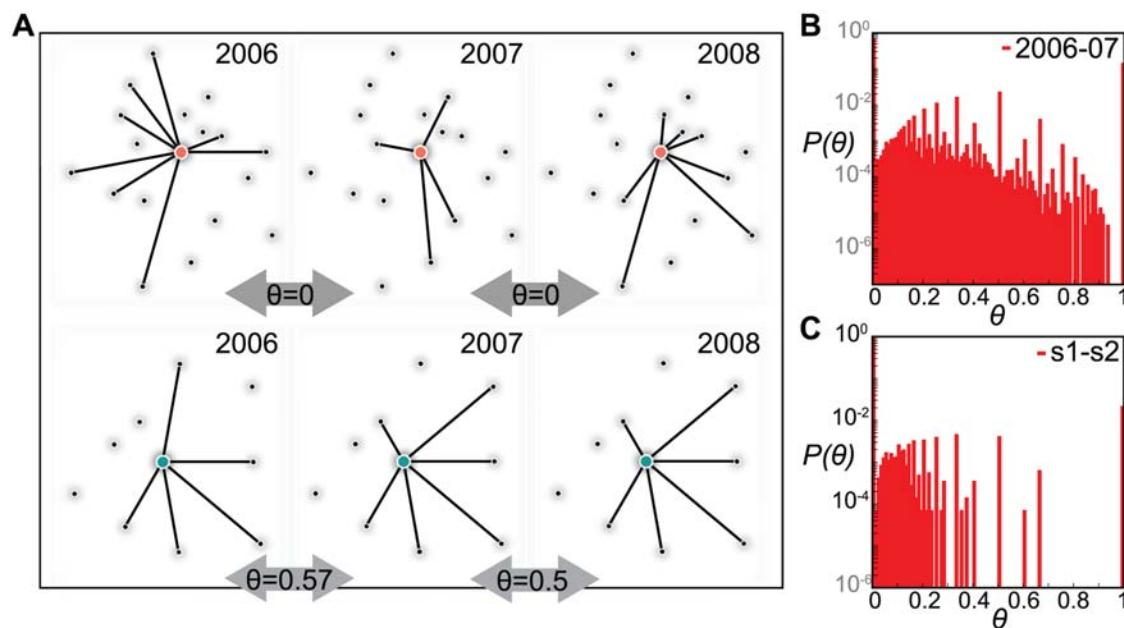
Fig 2. Loyalty. (*A*) Visualization of the neighborhood of two different farms in the cattle trade network (orange node, characterized by low loyalty, and green node, characterized by high loyalty) and corresponding loyalty values computed on three consecutive configurations (2006, 2007, 2008). (*B*), (*C*): Loyalty distributions in the cattle trade network and in the sexual contact network, respectively. Histograms refer to the first pair of consecutive configurations for visualization purposes, all other distributions being reported in S1 Text and showing stability across time.

doi:10.1371/journal.pcbi.1004152.g002

network topology and time-variation, for the prediction of the risk of infection. The aim is to provide a general and conceptually simple framework, leaving to future studies the investigation of more detailed and realistic disease natural histories.

At each time step, an infectious node can transmit the disease along its outgoing links to its neighboring susceptible nodes that become infected and can then propagate the disease further in the network. Here, we consider a deterministic process for which the contagion occurs with probability equal to 1, as long as there exist a link connecting the infectious node to a susceptible one. Although a crude assumption, this allows us to simplify the computational aspects while focusing on the risk prediction. The corresponding stochastic cases exploring lower probabilities of transmission per link are reported in S1 Text.

We focus on the early phase of the spreading simulations, defined as the set of nodes infected up to simulation time step $\tau = 6$. This choice allows us to study invasion stage only, while the epidemic is no more trivially confined to the microscopic level. Additional choices for τ have been investigated showing that they do not alter our findings (see S1 Text). Network configurations are kept constant during outbreaks, assuming diseases spread faster than network evolution, at least during their invasion stage. Examples of incidence curves obtained by the simulations are reported in S1 Text.

Livestock disease spread is often modeled by assuming that premises are the single discrete units of the spreading processes and neglecting the possible impact of within-farm dynamics [49]. This is generally considered in the study of highly contagious and rapid infections, and corresponds to regarding a farm as being infected as soon as it receives the infection from neighboring farms following the transport of contagious animals. Under this assumption, both case studies can be analyzed in terms of networks of contacts for disease transmission. In addition, for sake of simplicity, we do not take into account the natural definition of link weights

on cattle network, representing the size of the moved batches. In S1 Text we generalize our methodology to the weighted case, including a weighted definition of loyalty, reaching results similar to the unweighted case.

We consider an emerging epidemic unfolding on a network configuration c and starting from a single node (seed s), where the rest of the population of nodes is assumed to be initially susceptible. The details on the simulations are reported in the Material and Methods section. We define $\mathcal{I}_s^c$ the set of nodes infected during the early stage invasion. In order to explore how the network topology evolution alters the spread of the disease, we consider an outbreak unfolding on the previous configuration of the system, $c-1$, and characterized by the same epidemiological conditions (same epidemic parameters and same initial seed s). By comparing the set of infected nodes $\mathcal{I}_s^{c-1}$ obtained in configuration $c-1$ to $\mathcal{I}_s^c$, we can assess changes in the two sets and how these depend on the nodes' loyalty. We define a node's infection potential $\pi_L^{c-1,c}(s)$ ($\pi_D^{c-1,c}(s)$) measuring the probability that a node will be infected in configuration c by an epidemic starting from seed s, given that it was infected in configuration $c-1$ under the same epidemiological conditions and provided that its loyalty status is L (D):

$$\pi_L^{c-1,c}(s) \overset{def}{=} \text{Prob}\left[i \in \mathcal{I}_s^c \mid i \in \mathcal{I}_s^{c-1} \text{ and } i \in \{L\}\right],$$

$$\pi_D^{c-1,c}(s) \overset{def}{=} \text{Prob}\left[i \in \mathcal{I}_s^c \mid i \in \mathcal{I}_s^{c-1} \text{ and } i \in \{D\}\right],$$

where i is a node of the system. π_L and π_D thus quantify the effect of the temporal stability of the network at the local level (loyalty of a node) on the stability of a macroscopic process unfolding on the network (infection). They depend on the seed chosen for the start of the epidemic, on the pair ($c-1$, c) of network configurations considered along its evolution, and also on the threshold value ϵ assumed for the definition of the loyalty status of the nodes.

By exploring all seeds and computing the infection potentials for different couples of years, we obtain sharply peaked probability distributions of π_L and π_D around values that are well separated along the π axis. Results are qualitatively similar in both cases under study, with peaks reached for $\pi_L/\pi_D \simeq 2.5$ in the cattle trade network and $\pi_L/\pi_D \simeq 3$ in the sexual trade network (Fig. 3A-B). An observed infection in $c-1$, based on the knowledge of the epidemiological conditions and no information on the network evolution, is an indicator of an infection risk for the same epidemic in c more than twice larger for loyal farms with respect to disloyal farms. Analogously, loyal sex buyers have a threefold increase in their infection potential with respect to individuals having a larger turnover of partners. Remarkably, small values of loyalty threshold ϵ are able to correctly characterize the loyal behavior of nodes with status L. Results shown in Fig. 3A-B are obtained for $\epsilon = 0.1$. Findings are however robust against changes in the choice of the threshold value, as this is induced by the peculiar bimodal shape of the probability distribution curves for the loyalty (see S1 Text). This means that intermediate values of the local stability of the nodes (i.e. $\theta > \epsilon$) imply that a possible risk of being infected is strongly stable, regardless of the dynamics of the network evolution. Valid for all possible seeds and epidemiological conditions, this result indicates that the loyalty of a node can be used as an indicator for the node's risk of infection, which has important implication for the spreading predictability in case an outbreak emerges.

These results are obtained for temporally evolving networks where no further change induced by the epidemic is assumed to occur. Focusing on the initial stage of the outbreak, we disregard the effect of interventions (e.g. social distancing, quarantine of infectious nodes, movements bans) or of adaptive behavior following awareness [37, 50, 51, 52, 53, 54]. Such assumption relies on the study's focus on the initial stage of the epidemic that may be characterized by a silent spreading phase with propagation occurring before the alert or outbreak
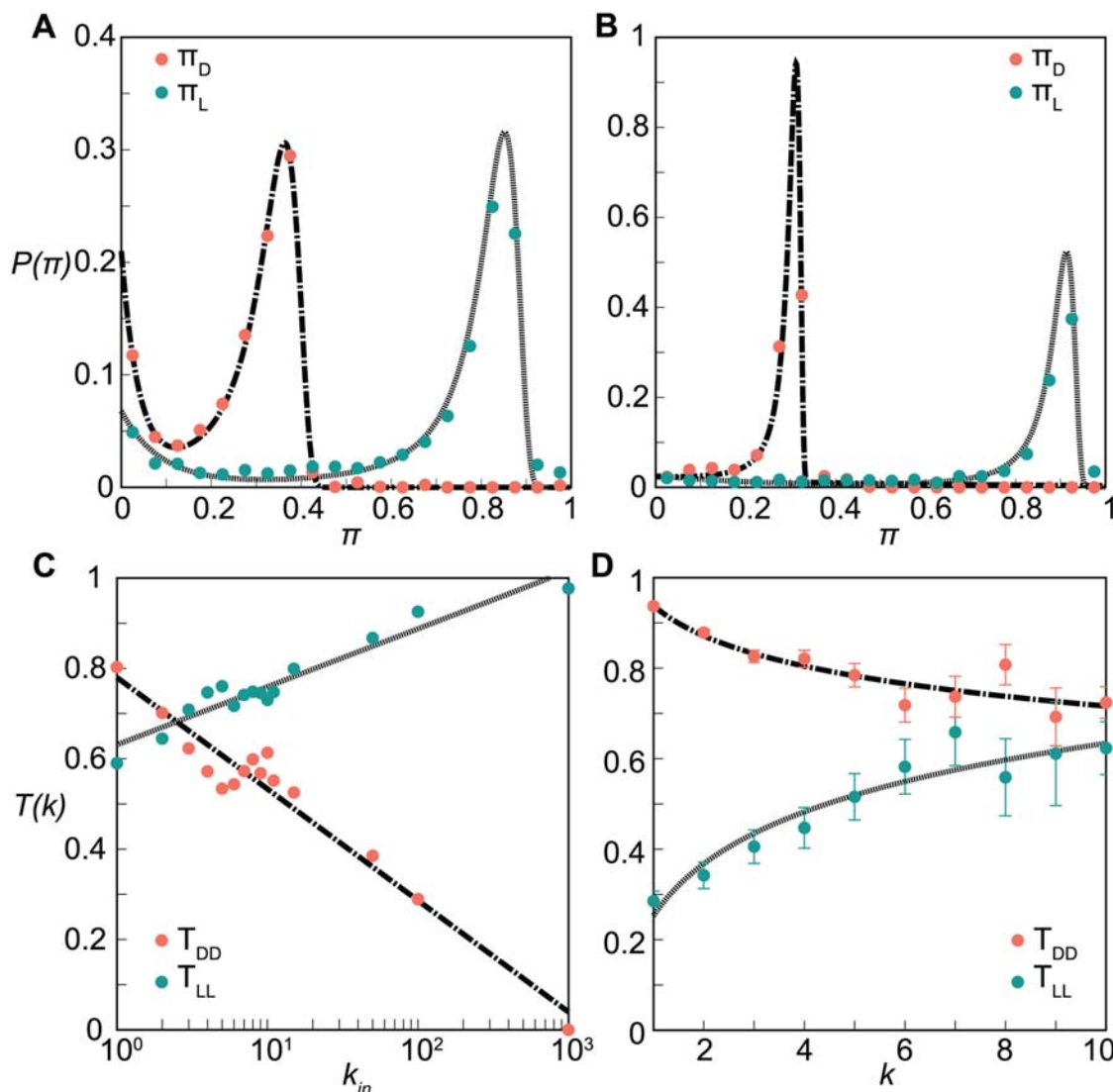
Fig 3. Infection potentials and loyalty transitions. (*A*), (*B*): Probability distributions of the infections potentials for loyal (π_L, green) and disloyal nodes (π_D, orange), for the cattle trade network and the sexual contact network, respectively. Loyalty is set with a threshold $\varepsilon = 0.1$. Dashed lines show the fit with a Landau+exponential model (see Material and [Methods](#)). (*C*), (*D*): Loyalty transition probabilities between loyal statuses ($T_{LL}(k)$, green) and disloyal statuses ($T_{DD}(k)$, orange) as functions of the degree k of the node, for the cattle trade network and the sexual contact network, respectively. Dashed lines represent the logarithmic models: $T_{DD}(k) = 0.78-0.11\log k$, and $T_{LL}(k) = 0.63+0.06\log k$ for the cattle trade network; $T_{DD}(k) = 0.94-0.10\log k$, and $T_{LL}(k) = 0.25+0.17\log k$ for the sexual contact network. Transition probabilities are computed as frequencies in the datasets under study. The error bars here represent one binomial standard deviation from these frequencies. In (*C*) the error bars are smaller than the size of the points. A single pair of configurations is considered here as example; the behavior observed is the same for all the pair of configurations.

doi:10.1371/journal.pcbi.1004152.g003

detection takes place; or, following an alert, by a contingent delay in the implementation of intervention measures.

Risk assessment analysis

The observed relationship between loyalty and infection potential can be used to define a strategy for the risk assessment analysis of an epidemic unfolding on an unknown networked system at present time, for which we have however information on its past configurations. This

may become very useful in practice even in the case of complete datasets, as for example with emerging outbreaks of livestock infectious diseases. Data on livestock movements are routinely collected following European regulations [55], however they may not be readily available in a real-time fashion upon an emergency, and a certain delay may thus be expected. Following an alert for an emerging livestock disease epidemic, knowledge of past network configurations may instead be promptly used in order to characterize the loyalty of farmers, simulate the spread of the disease on past configurations and thus provide the expected risk of infection for the farms under the ongoing outbreak. The general scheme of the strategy for the risk assessment analysis is composed of the following steps, assuming that the past network configurations $\{c-n, \ldots, c-1, c\}$ are known and that the epidemic unfolds on the unknown configuration $c+1$:

1. identify the seed s of the ongoing epidemic;

2. characterize the loyalty of the nodes from past configurations by computing $\theta_i^{c-1,c}$ from Eq. (1);

3. predict the loyalty of the nodes for the following unknown configuration $c+1$: $\theta_i^{c,c+1}$;

4. simulate the spread of the epidemic on the past configuration c under the same epidemiological conditions of the ongoing outbreak and identify the infected nodes $\mathcal{I}_s^c$;

5. compute the node epidemic risk for nodes in statuses L and D.

This strategy enables the assessment of the present infection risk (i.e. on configuration $c+1$) for all nodes hit by the simulated epidemic spreading on past configuration c ($\mathcal{I}_s^c$), not knowing their present pattern of contacts. It is based on configurations from $c-n$ to c as they are all used to build the probability distributions needed to train our approach. In the cases under study such distributions are quite stable over time so that a small set of configurations ($\{c-2, c-1, c\}$) was shown to be enough.

To make the above strategy operational, we still need to determine how we can exploit past data to predict the evolution of the loyalty of a node in future configurations (step 3) and use this information to compute nodes epidemic risk (point 5). As with all other variables characterizing the system, indeed, also θ may fluctuate from a pair of configurations ($c-1$, c) to another, as nodes may alter their loyal behavior over time, increasing or decreasing the memory of the system across time. Without any additional knowledge or prior assumption on the dynamics driving the system, we measure from available past data the probabilities of (dis)loyal nodes staying (dis)loyal across consecutive configurations, or conversely, of changing their loyalty status. This property can be quantified in terms of probabilities of transition across loyalty statuses. We thus define $T_{LL}^c(k)$ as the probability that a node with degree k being loyal between configurations $c-1$ and c will stay loyal one step after (c, $c+1$). It is important to note the explicit dependence on the degree k of the node (here defined at time c), which may increase or decrease following neighborhood reshaping (it may also assume the value $k = 0$ if the node becomes inactive in configuration c). Analogously, $T_{DD}^c(k)$ is the probability of remaining disloyal. The other two possible transition probabilities are easily obtained as $T_{LD} = 1-T_{LL}$ and $T_{DL} = 1-T_{DD}$.

Fig. 3C-D show the transition probabilities of maintaining the same loyalty status calculated on the two empirical networks for $\epsilon = 0.1$. Stability in time and non-trivial dependences on the degree of the node are found for both networks. In the cattle trade network, loyal farmers tend to remain loyal with a rather high probability ($T_{LL} > 0.6$ for all k_{in} values). In addition, this probability markedly increases with the degree, reaching $T_{LL} \simeq 1$ for the largest values of k_{in}.

Interestingly, the probability that a disloyal farmer stays disloyal the following year dramatically decreases with the degree, reaching 0 in the limit of large degree. Among the farmers who purchase cattle batches from a large number of different premises, loyal ones have an increased chance to establish business deals with the same partners the following year, whereas previously disloyal ones will more likely turn to being loyal.

A similar qualitative dependence on the degree is also found in the sexual contact network, however in this case the probability of remaining disloyal is always very high ($T_{DD} > 0.7$) even for high degrees. T_{LL} shows a relatively more pronounced dependence on k, ranging from 0.3 (low degree nodes) to 0.6 (high degree nodes). Differently from the farmers behavior, sex buyers display a large tendency to keep a high rate of partners turnover across time. Moreover, the largest probability of preserving sexual partners is obtained when the number of partners is rather large.

Remarkably, in both networks, transition probabilities are found to be stable across time and are well described by logarithmic functions (with parameters depending on the system and on ϵ) that can be used to predict the loyalty of nodes in configuration $c+1$ from past data (Fig. 3C-D). With this information, it is then possible to compute the epidemic risk of a node i in configuration $c+1$, having degree $k = k_i^c$ in configuration c and known loyalty status $\{L, D\}$ between configurations $c-1$ and c as follows:

$$
\begin{cases}
\text{if loyalty class} = D: & \rho_i^{c+1} = \pi_D^{c,c+1}(s)T_{DD}(k) + \pi_L^{c,c+1}(s)T_{DL}(k) ; \\
\text{if loyalty class} = L: & \rho_i^{c+1} = \pi_D^{c,c+1}(s)T_{LD}(k) + \pi_L^{c,c+1}(s)T_{LL}(k) .
\end{cases}
\tag{2}
$$

It is important to note that in our framework the epidemic risk is a node property, and not a global characteristic of a specific disease.

Validation

To validate our strategy of risk assessment, we test our predictions based on past data for the risk of being infected in configuration $c+1$ on the results of an epidemic simulation explicitly performed on the supposedly unknown configuration $c+1$. We consider the set of nodes $\mathcal{I}_s^c$ for which we are able to provide risk predictions and divide it into two subsets, according to their predicted risk of infection ρ_i^{c+1}. We indicate with $\mathcal{I}_{s,h}^c$ the top 25% highest ranking nodes, and with $\mathcal{I}_{s,l}^c$ all the remaining others. We then compute the fraction P_h of nodes in the subset $\mathcal{I}_{s,h}^c$, i.e. predicted at high risk, that belong to the set of infected nodes $\mathcal{I}_s^{c+1}$ in the simulated epidemic aimed at validation. Analogously, P_l measures the fraction of nodes in $\mathcal{I}_{s,l}^c$ that are reached by the infection in the simulation on $c+1$. In other words, P_h (P_l) represents the probability for a node having a high (low) risk of infection to indeed get infected. The accuracy of the risk assessment analysis can thus be measured in terms of the relative risk ratio $v = P_h/P_l$, where values $v \leq 1$ indicate negative or no correlation between our risk predictions and the observed infections, whereas values $v > 1$ indicate that the prediction is informative. For both networks we find a significant correlation, signaled by the distributions of the relative risk ratio v peaking around values $v > 1$ (Fig. 4A-B). The peak positions ($v \simeq 1.4$ and $v \simeq 1.7$ for cattle and sex, respectively) are remarkably close to the benchmark values represented by the distributions computed on the training sets (red lines in Fig. 4A-B). In addition, the comparison with the distributions from a null model obtained by reshuffling the infection statuses of nodes (dotted curves peaking around $v = 1$ in Fig. 4A-B) further confirms the accuracy of the approach. Findings are robust against changes of the value used to define $\mathcal{I}_{s,h}^c$ or against alternative definitions of this quantity (see S1 Text).
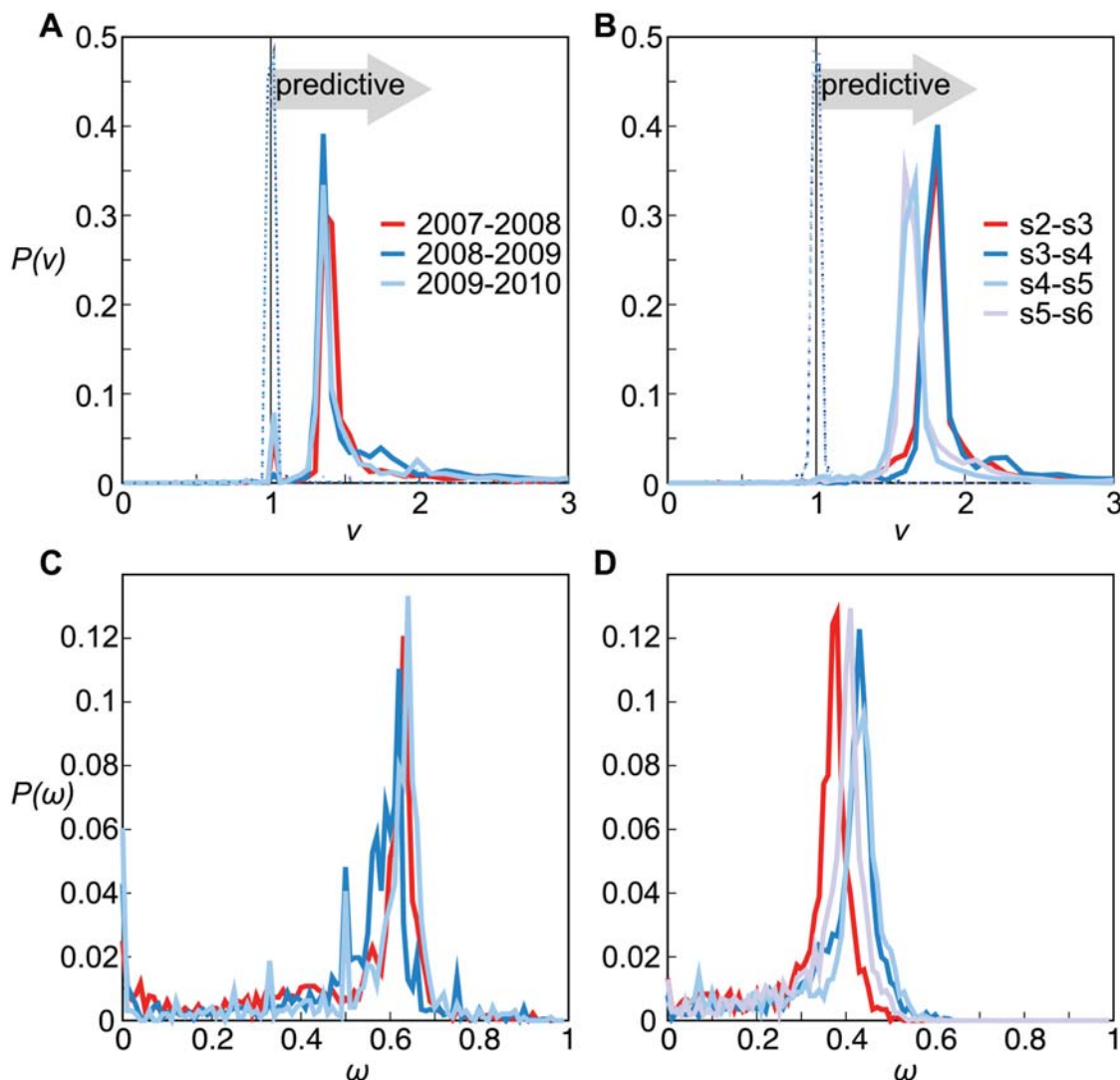
Fig 4. Validation of the risk assessment analysis. (A), (B): Probability distributions of the risk ratio v for the cattle trade network and the sexual contact network, respectively. Red lines are computed on training sets (2007–08 for cattle and s2-s3 for sexual contacts). The dashed lines peaking around 1 represent a null model based on reshuffling the infection statuses, i.e. randomly permuting the attribute *"actually being infected"* among the nodes for which risk assessment is performed. (C), (D): Probability distributions of the predictive power ω for the cattle trade network and the sexual contact network, respectively.

One other important aspect to characterize is the predictive power of our risk assessment analysis. Our predictions indeed are limited to the set $\mathcal{I}_s^c$ of nodes that are reached in the simulation performed on past data, proxy for the future outbreak. If a node is not infected by the simulation unfolding on configuration c or it is not active at that given time, our strategy is unable to provide a risk assessment for that node in the future. We can then quantify the predictive power ω as the fraction of infected nodes for which we could provide the epidemic risk, i.e. $\omega_s^{c,c+1} = |\mathcal{I}_s^{c+1} \cap \mathcal{I}_s^c| / |\mathcal{I}_s^{c+1}|$. High values of ω indicate that few infections are missed by the risk assessment analysis. Fig. 4C-D display the distributions $P(\omega)$ obtained for the two case studies, showing that a higher predictive power is obtained in the cattle trade network (peak at $\omega \simeq 60\%$) with respect to the sexual contact network (peak at $\omega \simeq 40\%$). Our methodology can

potentially be applied to a wide range of networks, other than the ones presented here, as shown with the example of human face-to-face proximity networks relevant for the spread of respiratory diseases reported in S1 Text.

We also tested whether our risk measure represents a significant improvement in prediction accuracy with respect to simpler and more immediate centrality measures (namely, the degree). Through a multivariate logistic regression, in S1 Text we show that our definition of node risk is predictor of infection even after adjusting for node degree.

Memory driven dynamical model

The results of the risk assessment analysis obtained from the application of our strategy to the two empirical networks show qualitatively similar results, indicating that the approach is general enough to provide valuable information on the risk of infection in different settings. The observed differences in the predictive power of the approach are expected to be induced by the different temporal behavior of the two systems, resulting in a different amount of memory in preserving links (Fig. 1) and different loyalty of nodes and their time-variations (Fig. 2 and 3C-D).

In order to systematically explore the role of these temporal features on the accuracy and predictive power of our approach, we introduce a generic model for the generation of synthetic temporal networks. The model is based on a set of parameters that can be tuned to reproduce the empirically observed features of the two networks, i.e.: *(i)* the topological heterogeneity of each configuration of the network described by a stable probability distribution (Fig. 1A-B); *(ii)* a vital dynamics to allow for the appearance and disappearance of nodes; *(iii)* a tunable amount of memory characterizing the time evolution of the network contacts (Fig. 1C-D). These specific properties differentiate our approach from the previously introduced models that display instantaneous homogeneous properties for network configurations [56, 57, 58, 59], reproduce bursty inter-event time distributions but without the explicit introduction of memory [33, 60, 61] or of its control [58].

Based on an iterative network generation approach (see Materials and Methods), we can build an arbitrarily large number of configurations of networks with 10^4 nodes. They are characterized by stable in-degree and out-degree heterogeneous distribution across time (Fig. 5A where high memory and low memory regimes are displayed) and by profiles for the probability distribution of the loyalty as in the empirical networks (Fig. 5B). The number of nodes with zero loyalty can be computed analytically (see Materials and Methods) and it is confirmed by numerical findings (see S1 Text). A high memory regime corresponds to having nodes in the system that display a highly loyal behavior (e.g., $\theta > 0.7$), whereas values in the range $\theta \in [0.7, 1)$ are almost absent in a low memory regime, in agreement with the findings of Fig. 2.

Applying the introduced risk assessment analysis to the synthetically generated temporal network, we recover a significant accuracy for both memory regimes (Fig. 5C). Different degrees of memory are however responsible for the fraction of the system for which a risk assessment can be made. In networks characterized by higher memory, the distribution of the predictive power ω has a well defined peak, whereas for lower memory it is roughly uniform in the range $\omega \in [0, 0.4]$ (Fig. 5D). Such a regime implies that not enough structure is maintained in the system to control more than 40% of the future infections. Our risk assessment analysis allows therefore accurate predictions across varying memory regimes characterizing the temporal networks, but the degree of memory impacts the amount of predictions that can be made. The model also shows that the analysis is not affected by the choice of the aggregating time window used to define the network configurations [61, 62, 63], as long as the heterogeneous topological features at the system level and the heterogeneous memory at the node level
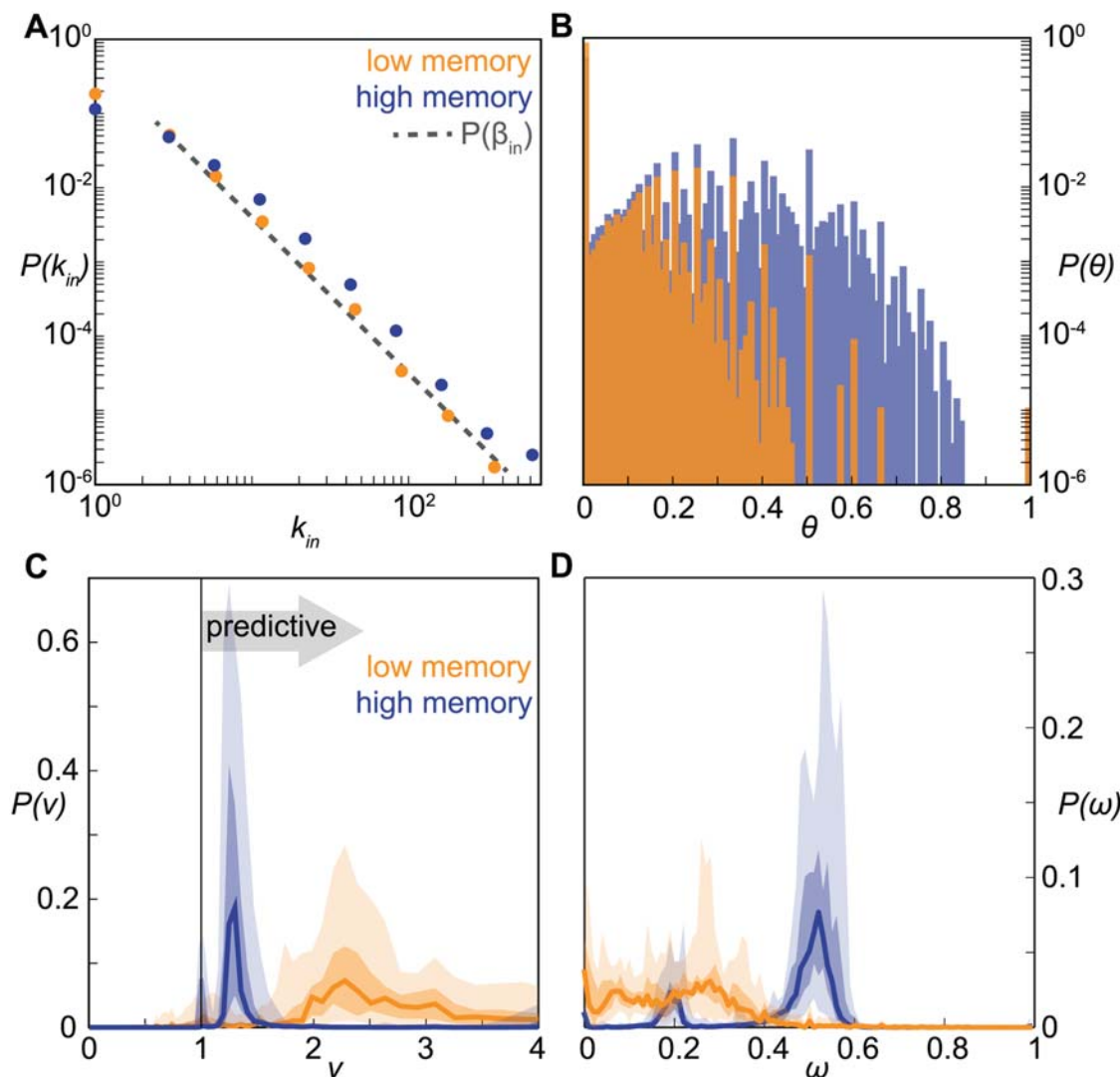
Fig 5. Memory driven dynamical model: model properties and validation of the risk assessment analysis. (A): Probability distributions of the node in-degree, in the low ($p_\alpha = 0.3$) and high memory ($p_\alpha = 0.7$) regimes. The slope of the distributions does not depend on p_α, and it is forced by the exponent γ of the β_{in} distribution (dashed line). (B): Probability distributions of the loyalty, in the low and high memory regimes. Distributions are color-coded as in panel (a). (C): Probability distributions of the risk ratio v, in the low and high memory regimes. Lines represent the median values obtained from 50 realizations of the model; darker and lighter shaded areas represent the 50% and 95% confidence intervals. (D): Probability distributions of the predictive power ω, in the low and high memory regimes. Medians and confidence intervals are presented as in panel (C). Distributions are color-coded as in panel (A).

doi:10.1371/journal.pcbi.1004152.g005

are kept across aggregation, as observed for the empirical networks under study (see [19] and S1 Text).

Conclusions

We introduce a simple measure to characterize the amount of memory in the time evolution of a networked system. The measure is local and it is empirically motivated from two case studies relevant for disease transmission. By focusing on the degree of loyalty that each node has in establishing connections with the same partners as time evolves, we are able to connect an ego-centric view of the system (the node's strategy in establishing its neighborhood over time) to

the system's larger-scale properties characterizing the early propagation of an emerging epidemic.

We uncover a non-trivial correlation between the loyalty of a node and its risk of being infected if an epidemic occurs, given fixed epidemiological conditions, and use this to inform a risk assessment analysis applicable to different settings with no information on the network evolution dynamics. A theoretical model generating synthetic time-varying networks allows us to frame the analysis in a more general perspective and disentangle the role of different features. The accuracy of the proposed risk assessment analysis is stable across variations of the temporal correlations of the system, whereas its predictive power depends on the degree of memory kept in the time evolution. The introduced strategy can be used to inform preventive actions in preparation to an epidemic and for targeted control responses during an outbreak emergency, only relying on past network data.

Methods

Datasets

The cattle trade network is obtained from the database of the Italian national bovine registry recording all cattle displacements due to trade transactions. We consider animal movements during a 5 years time period, from 2006 to 2010, involving 215,264 premises and 2,973,710 directed links. Nodes may be active or inactive depending whether farms sell/buy cattle in a given timeframe. The cattle network is available as S1 Dataset. From the dataset we have removed slaughterhouses ($\sim$ 1% of the nodes) as they are not relevant for transmission.

The sexual contact network is extracted from an online Brazilian forum where male sex buyers rate and comment on their sexual encounters with female sex sellers [16]. Time-stamped posts are used as proxies for sexual intercourse and multiple entries are considered separately, following previous works [16, 31]. A total of 13,855 individuals establishing 34,509 distinct sexual contacts are considered in the study, after discarding the initial transient of the community growth [16]. Nodes may be active or inactive depending whether individuals use or not the service, and join or quit the community. Six-months aggregating snapshots are chosen. A different aggregating time window of three months has been tested, obtaining similar results (see S1 Text).

Risk of infection

The distributions of the risk potentials π_L and π_D reported in Fig. 3 are modeled with a sum of Landau distribution and an exponential suppression. This family of functions depending on four parameters (see S1 Text for the specific functional form) was chosen as it well reproduces the distribution profiles of the risk potentials, and it was used to compute the nodes' epidemic risk. A goodness of fit was not performed, as this choice was automatically validated in the validation analysis performed on the whole prediction approach.

Memory driven model

The basic iterative network generation approach allows to build configuration c+1 from configuration c through the following steps:

- vital dynamics: nodes that are inactive in configuration c become active in c+1 with probability b, while active nodes become inactive with probability d;

- memory: active nodes maintain same in-neighbors each with probability p_α; then they form β_{in} new in-stubs, where β_{in} is extracted from a power-law distribution: $P(\beta_{in}) \sim \beta_{in}^{-\gamma}$;

- out-degree heterogeneity: each node is assigned β_{out} out-stubs, where β_{out} is drawn from another power-law distribution: $P(\beta_{out}) \sim \beta_{out}^{-\delta}$. Then each of the in-stubs is randomly matched to an out-stub.

The total set of nodes is considered to be fixed in time, and nodes may be active (i.e. establishing connections) or inactive (i.e. isolated) in a given configuration. All five parameters b, d, γ, p_α, δ are assumed constant in time and throughout the network. The amount of memory in the system is tuned by the interplay of the two parameters p_α and d. Starting from an arbitrarily chosen initial configuration $c = 0$, simulations show that the system rapidly evolves towards a dynamical equilibrium, and successive configurations can be obtained after discarding an initial transient of time. The parameters values used in the paper are: $N = 10^4$; $b = 0.7$; $d = 0.2$; $\gamma = 2.25$; $\delta = 2.75$; $p_\alpha = 0.3, 0.7$. The influence of such parameters on the network properties is examined in S1 Text.

If we denote with α the number of neighbors that a given node keeps across two consecutive configurations $(c-1, c)$, we can express the loyalty simply as:

$$\theta_i^{c-1,c} = \frac{\alpha^{c-1,c}}{\left(k_i^c + \beta_{in}^c\right)} \tag{3}$$

where the superscript c for α, β_{in} indicate the values used to build configuration c. The number of nodes with $\theta = 0$ as a function of the degree can be computed analytically: $P(\theta_{c,c+1} = 0) = d + (1 - d)(1 - p_\alpha)^{k_c}$. Similarly, it is possible to compute the probability $f_{c, c+1}$ that a link present in configuration c is also present in configuration $c+1$. In the S1 Text we show that $f_{c, c+1} \simeq (1-d)p_\alpha$ and confirm this result by numerical simulations.

Supporting Information

S1 Dataset. Cattle trade network dataset. We provide the cattle trade network as yearly edge lists, from 2006 to 2010. The dataset consists in five CSV files (one for each year) compressed in a ZIP archive.
(ZIP)

S1 Text. Additional analyses. We provide a description of the seasonal pattern of cattle network (Section 1), a more in-depth characterization of loyalty (Section 2), a comparison between loyalty and other similarity measures (Section 3), the specific modeling function for the infection potential (Section 4), the robustness of the risk assessment procedure to variations in parameters and assumptions (Section 5), further analyses of the memory driven model in terms of analytical results (Section 6) and additional properties (Section 7), an extension of our methodology to take into account transmissibility lower than 1 (Section 8), and links weights (Section 9).
(PDF)

Author Contributions

Conceived and designed the experiments: EV CP VC. Performed the experiments: EV. Analyzed the data: EV CP VC. Contributed reagents/materials/analysis tools: EV CP AG DP LS VC. Wrote the paper: EV CP AG DP LS VC.

References

1. Pastor-Satorras R, Vespignani A (2001) Epidemic spreading in scale-free networks. Phys Rev Lett 86:3200–3203. doi: 10.1103/PhysRevLett.86.3200 PMID: 11290142

2. Lloyd A, May R (2001) Epidemiology: how viruses spread among computers and people. Science 292:1316–1317. doi: 10.1126/science.1061076 PMID: 11360990

3. Newman MEJ (2002) Spread of epidemic disease on networks. Phys Rev E 66, 016128. doi: 10.1103/PhysRevE.66.016128

4. Cohen R, Havlin S, ben-Avraham D (2003) Efficient immunization strategies for computer networks and populations. Phys Rev Lett 91:247901. doi: 10.1103/PhysRevLett.91.247901 PMID: 14683159

5. Keeling MJ, Eames KTD (2005) Networks and epidemic models. J R Soc Interface 2: 295–307. doi: 10.1098/rsif.2005.0051 PMID: 16849187

6. Riley S (2007) Large-scale spatial-transmission models of infectious disease. Science 316: 1298–1301. doi: 10.1126/science.1134695 PMID: 17540894

7. Colizza V, Barrat A, Barthélémy M, Vespignani A (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. Proc Natl Acad Sci USA 103:2015–2020. doi: 10.1073/pnas.0510525103 PMID: 16461461

8. Brockmann D, Helbing D (2013) The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. Science 342, 1337. doi: 10.1126/science.1245200 PMID: 24337289

9. Riley S et al (2003) Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions. Science 300: 1961–1966. doi: 10.1126/science.1086478 PMID: 12766206

10. Eckmann JP, Moses E, Sergi D (2004) Entropy of dialogues creates coherent structures in e-mail traffic. Proc Natl Acad Sci USA 101:14333–14337. doi: 10.1073/pnas.0405728101 PMID: 15448210

11. Onnela JP, Saramäki J, Hyvönen J, Szabó G, Lazer D, et al. (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci USA 104(18):7332–7336. doi: 10.1073/pnas.0610245104 PMID: 17456605

12. Rybski D, Buldyrev SV, Havlin S, Liljeros F, Makse HA (2009) Scaling laws of human interaction activity. Proc Natl Acad Sci USA 106(31):12640–12645. doi: 10.1073/pnas.0902667106 PMID: 19617555

13. Gautreau A, Barrat A, Barthélémy M (2009) Microdynamics in stationary complex networks. Proc Natl Acad Sci (USA) 106:8847–8852. doi: 10.1073/pnas.0811113106

14. Vernon MC, Keeling MJ (2009) Representing the UK's cattle herd as static and dynamic networks. Proc R Soc B 276:469–476 doi: 10.1098/rspb.2008.1009 PMID: 18854300

15. Cattuto C, Van den Broek W, Barrat A, Colizza V, Pinton JF, et al. (2010) Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. PLOS ONE 5(7):e11596. doi: 10.1371/journal.pone.0011596 PMID: 20657651

16. Rocha LEC, Liljeros F, and Holme P (2010) Information dynamics shape the sexual networks of Internet-mediated prostitution. Proc Natl Acad Sci USA 107:5706–5711. doi: 10.1073/pnas.0914080107 PMID: 20231480

17. Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, et al. (2010) A high-resolution human contact network for infectious disease transmission. Proc Natl Acad Sci USA 107:22020–22025. doi: 10.1073/pnas.1009094108 PMID: 21149721

18. Tang J, Scellato S, Musolesi M, Mascolo C, Latora V (2010) Small-world behavior in time-varying graphs. Phys Rev E 81:055101. doi: 10.1103/PhysRevE.81.055101

19. Bajardi P, Barrat A, Natale F, Savini L, Colizza V (2011) Dynamical patterns of cattle trade movements. PLOS ONE, 6(5):e19869 doi: 10.1371/journal.pone.0019869 PMID: 21625633

20. Miritello G, Moro E, Lara R (2011) Dynamical strength of social ties in information spreading. Phys Rev E 83:045102. doi: 10.1103/PhysRevE.83.045102

21. Karsai M, Kaski K, Barabási AL, Kertész J (2012) Universal features of correlated bursty behaviour. Sci Rep 2, 397. doi: 10.1038/srep00397 PMID: 22563526

22. Holme P, Saramaki J (2012) Temporal networks. Phys Rep 519:97. doi: 10.1016/j.physrep.2012.03.001

23. Albert R, Jeong H, Barabási AL (2000) Error and attack tolerance of complex networks. Nature 406:378–482. doi: 10.1038/35019019 PMID: 10935628

24. Cohen R, Erez K, ben-Avraham D, Havlin S (2001) Breakdown of the Internet under intentional attack. Phys Rev Lett 86:3682–3685. doi: 10.1103/PhysRevLett.86.3682 PMID: 11328053

25. Freeman LC (1979) Centrality in social networks: Conceptual clarification. Social Networks 1:215–239. doi: 10.1016/0378-8733(78)90021-7

26. Friedkin NE (1991) Theoretical foundations for centrality measures. Am J Sociology 96:1478–1504. doi: 10.1086/229694

27. Holme P (2004) Efficient local strategies for vaccination and network attack. Europhys Lett 68:908–914. doi: 10.1209/epl/i2004-10286-2

28. Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, et al. (2010) Identification of inuential spreaders in complex networks. Nature Phys 6:888–893. doi: 10.1038/nphys1746

29. Salathé M, Jones JH (2010) Dynamics and control of diseases in networks with community structure. PLOS Comput Biol 6:e1000736. doi: 10.1371/journal.pcbi.1000736 PMID: 20386735

30. Natale F, Savini L, Giovannini A, Calistri P, Candeloro L, et al. (2011) Evaluation of risk and vulnerability using a Disease Flow Centrality measure in dynamic cattle trade networks. Prev Vet Med 98(2–3):111–8. doi: 10.1016/j.prevetmed.2010.11.013 PMID: 21159393

31. Rocha LEC, Liljeros F, Holme P (2011) Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts. PLOS Computational Biology 7(3) (2011) doi: 10.1371/journal.pcbi.1001109 PMID: 21445228

32. Galvani AP, May RM (2005) Dimensions of superpsreading. Nature 438, 293–294. doi: 10.1038/438293a PMID: 16292292

33. Lee S, Rocha LEC, Liljeros F, Holme P (2012) Exploiting temporal networks structures of human interaction to effectively immunize populations. PLOS ONE 7(5):e36439. doi: 10.1371/journal.pone.0036439 PMID: 22586472

34. Starnini M, Machens A, Cattuto C, Barrat A, Pastor-Satorras R (2013) Immunization strategies for epidemic processes in time-varying contact networks. J Theor Biol 337: 89–100. doi: 10.1016/j.jtbi.2013.07.004 PMID: 23871715

35. Bajardi P, Barrat A, Savini L, Colizza V (2012) Optimizing surveillance for livestock disease spreading through animal movements. J Roy Soc Int (June, 2012). doi: 10.1098/rsif.2012.0289

36. Natale F, Giovannini A, Savini L, Palma D, Possenti L, et al. (2009) Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. Prev Vet Med 92:341–350. doi: 10.1016/j.prevetmed.2009.08.026 PMID: 19775765

37. Robinson SE, Everett MG, Christley RM (2007) Recent network evolution increases the potential for large epidemics in the British cattle population. J Roy Soc Int 4: 669–674 doi: 10.1098/rsif.2007.0214

38. Kao RR, Danon L, Green DM, Kiss IZ (2006) Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. Proc Roy Soc Lond B Biol Sci 273: 1999–2007 doi: 10.1098/rspb.2006.3505

39. Sharp B, Sharp A (1997) Loyalty Programs and Their on Repeat-Purchase Loyalty Patterns. International Journal of Research in Marketing 14(5):473–86. doi: 10.1016/S0167-8116(97)00022-0

40. Podolny JM (1994) Market uncertainty and the social character of economic exchange. Administrative Science Quarterly 39:458–483. doi: 10.2307/2393299

41. Sorenson O, Waguespack DM (2006) Structure and Exchange: Self-Confirming Dynamics in Hollywood. Administrative Science Quarterly 51:560–589.

42. Miritello G, Lara R, Cebrian M, Moro E (2013) Limited communication capacity unveils strategies for human interaction. Sci Rep 3, 1950. doi: 10.1038/srep01950 PMID: 23739519

43. Morris M, Kretzschmar M (1997) Concurrent partnerships and the spread of HIV. AIDS 11:641–648. doi: 10.1097/00002030-199705000-00012 PMID: 9108946

44. Anderson RM, May RM, Boily MC, Garnett GP, Rowley JT (1991) The spread of HIV-1 in Africa: sexual contact patterns and the predicted demographic impact of AIDS. Nature 352:581–589. doi: 10.1038/352581a0 PMID: 1865922

45. Keeling MJ, Woolhouse MEJ, Shaw DJ, Matthews L, Chase-Topping M, et al. (2001) Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. Science 294:813–817 doi: 10.1126/science.1065973 PMID: 11679661

46. Saegerman C, Berkvens D, Mellor PS (2008) Bluetongue Epidemiology in the European Union. Emerging Inf Dis 14:539–544 doi: 10.3201/eid1404.071441

47. Tinsley M, Lewis FI, Brülisauer F (2012) Network modeling of BVD transmission Veterinary Research 43:11

48. Anderson RM, May RM (1992) Infectious Diseases of Humans Dynamics and Control. Oxford University Press, Oxford.

49. Keeling MJ (2005) Models of foot-and-mouth disease. Proc Roy Soc B 272:1195–1202. doi: 10.1098/rspb.2004.3046

50. Gross T, Dommar D'Lima CJ, Blasius B (2006) Epidemic dynamics on adaptive networks. Phys Rev Lett 96:208701. doi: 10.1103/PhysRevLett.96.208701 PMID: 16803215

51. Funk S, Salathé M, Jensen VAA (2010) Modeling the inuence of human behavior on the spread of infectious diseases: a review. J R Soc Interface 7:1247–1256. doi: 10.1098/rsif.2010.0142 PMID: 20504800

52. Shaw LB, Schwartz IB (2010) Enhanced vaccine control of epidemics in adaptive networks. Phys Rev E 81:046120. doi: 10.1103/PhysRevE.81.046120

53. Meloni S, Perra N, Arenas A, Gomez S, Moreno Y, et al. (2011) Modeling human mobility responses to the large-scale spreading of infectious diseases. Sci Rep 1:62. doi: 10.1038/srep00062 PMID: 22355581

54. Bajardi P, Poletto C, Ramasco JJ, Tizzoni M, Colizza V, et al. (2011) Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. PLOS ONE 6(1), e16591. doi: 10.1371/journal.pone.0016591 PMID: 21304943

55. European Parliament and European council (2000) 204: 1–10. Regulation (EC) No. 1760/2000 of 17 July 2000 establishing a system for the identification and registration of bovine animals and regarding labeling of beef and beef products and repealing Council Regulation (EC) No. 820/97 European Council. Off. J. Eur.Communities L.

56. Stehlé J, Barrat A, Bianconi G (2010) Dynamical and bursty interactions in social networks. Phys. Rev. E 81:035101. doi: 10.1103/PhysRevE.81.035101

57. Perra N, Gonçalves B, Pastor-Satorras R, Vespignani A (2012) Activity driven modeling of time varying networks. Sci Rep 2:469. doi: 10.1038/srep00469 PMID: 22741058

58. Starnini M, Baronchelli A, Pastor-Satorras R (2013) Modeling Human Dynamics of Face-to-Face Interaction Networks. Phys Rev Lett 110:168701 doi: 10.1103/PhysRevLett.110.168701 PMID: 23679648

59. Kársai M, Perra N, Vespignani A (2014) Time varying networks and the weakness of strong ties. Sci Rep 4:4001. doi: 10.1038/srep04001 PMID: 24510159

60. Rocha LEC, Blondel VD (2013) Bursts of Vertex Activation and Epidemics in Evolving Networks. PLoS Comput Biol 9(3): e1002974. doi: 10.1371/journal.pcbi.1002974 PMID: 23555211

61. Holme P (2013) Epidemiologically optimal static networks from temporal network data. PLOS Comp Bio. doi: 10.1371/journal.pcbi.1003142

62. Hoffmann T, Porter MA, Lambiotte R (2012) Generalized master equations for non-Poisson dynamics on networks. Phys Rev E 86:046102. doi: 10.1103/PhysRevE.86.046102

63. Ribeiro B, Perra N, Baronchelli A (2013) Quantifying the effect of temporal resolution on time-varying networks. Sci Rep 3:3006. doi: 10.1038/srep03006 PMID: 24141695

Predicting epidemic risk from past temporal contact data

Eugenio Valdano[a,b], Chiara Poletto[a,b], Armando Giovannini[c], Diana Palma[c],
Lara Savini[c], and Vittoria Colizza[a,b,d]

[a]INSERM, UMR-S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013,
Paris, France. [b]Sorbonne Universités, UPMC Univ Paris 06, UMR-S 1136, Institut Pierre Louis
d'Epidémiologie et de Santé Publique, F-75013, Paris, France. [c]Istituto Zooprofilattico
Sperimentale Abruzzo-Molise G. Caporale, Teramo, Italy. [d]ISI Foundation, Torino, Italy.

1 Seasonal pattern in cattle trade network

Fig. S1 shows the number of active links per month in the cattle trade network. A
seasonal pattern is clearly visible: the activity drops during summer months, and
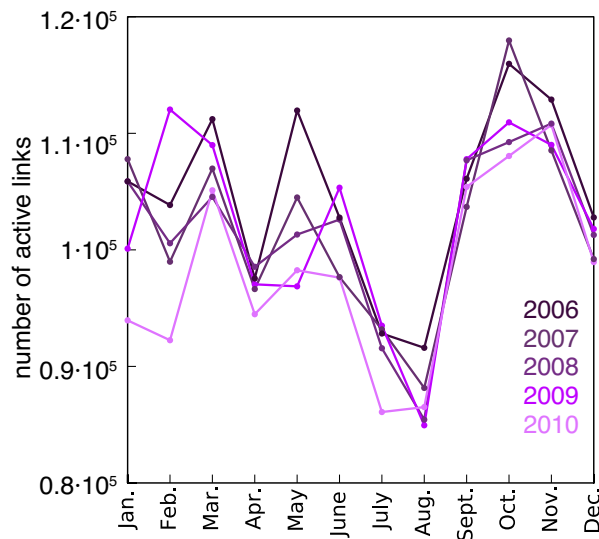peaks during fall. The activity pattern is quite similar from one year to the other.

Figure S1: Number of active links per month in cattle trade network. Different colors
pertain to different years, in range $(2006 - 2010)$.

2 Loyalty's properties

2.1 Allowed values

In the following we provide an analytical reasoning on the allowed values for the loyalty. θ between configurations c and $c + 1$ can be rewritten as

$$\theta = \frac{\alpha}{A - \alpha}, \tag{1}$$

where $\alpha \in \mathbb{N}$ is the number of neighbors retained from c to $c+1$, and $A = k_c + k_{c+1} \in \mathbb{N}$ is the sum of the node's degrees. Clearly, every pair of α', A' for which $\exists q \in \mathbb{N}$ such that $\alpha' = q\alpha$ and $A' = qA$, will give the same θ. Therefore, in order to compute all the possible values of θ, we must restrict ourselves to α, A coprimes: $(\alpha, A) = 1$. Moreover, since θ cannot be higher than 1, we have to impose one further constraint: $\alpha < A/2$. All divisions are to be intended as integer divisions.

For zero loyalty, we have $\theta = 0 \Leftrightarrow \alpha = 0$, for every positive A. If $\theta > 0$, we need to count the number of possible values α, given the constraints discussed above, and given a value for A which is fixed by the node's degrees. For $A \geq 3$, there are $\varphi(A)/2$ coprimes of A and smaller than or equal to $A/2$, as it can be inferred by basic properties of the Euler's totient function φ.

$$n(A) = \begin{cases} 0 & \text{if } A = 1 \\ 1 & \text{if } A = 2 \\ \varphi(A)/2 & \text{if } A \geq 3 \end{cases}, \tag{2}$$

where $n(A)$ counts the number of nonzero allowed values for θ, given a fixed A. In order to compute the total number of allowed θ values in an entire network, we now let A run from 1 to a certain A_{max}, which is of the order of twice the highest degree:

$$\mathcal{N}(A_{max}) = 1 + \sum_{A=1}^{A_{max}} n(A) = 2 + \frac{1}{2} \sum_{A=3}^{A_{max}} \varphi(A). \tag{3}$$

The unity added to the sum takes into account the value $\theta = 0$. In order to better understand the behavior of $\mathcal{N}(A_{max})$ we can use Walfisz approximation for large A_{max}, and assume $A_{max} \approx 2k_{max}$ to get

$$\mathcal{N}(k_{max}) = 1 + \frac{6}{\pi^2} k_{max}^2 + \mathcal{O}\left[k_{max} \left(\log k_{max}\right)^{2/3} \left(\log \log k_{max}\right)^{4/3}\right]. \tag{4}$$

This means that the sexual contact network has $\sim 10^4$ allowed values, and the cattle trade network has $\sim 10^8$ allowed values. Such large number of allowed values in the interval $[0, 1]$ justifies our approximation of treating θ as a continuous variable.

2.2 Temporal stability of the loyalty distribution in cattle and sexual contact networks

Fig. S2 shows the loyalty distributions in all configuration pairs included in the two datasets under study (top, cattle trade network; bottom, sexual contact network). In both networks, distributions are stable in time.
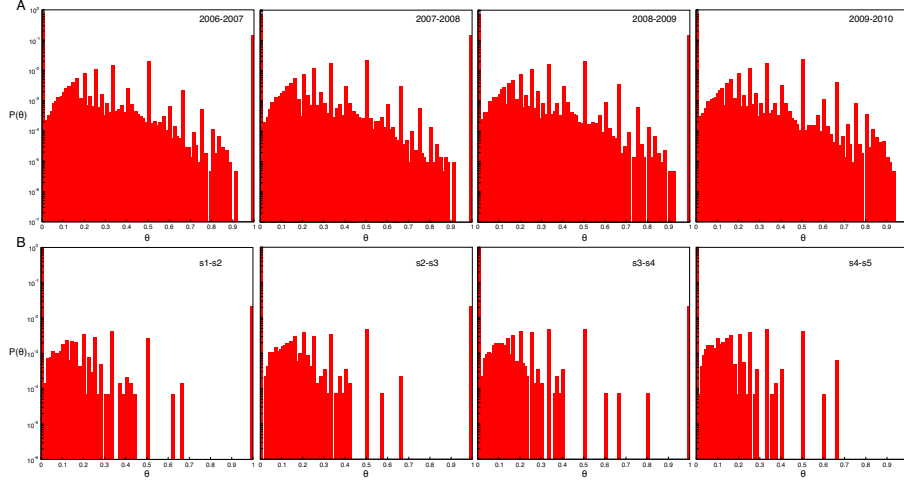
Figure S2: Loyalty distributions for different configurations. (*A*): distributions for cattle network, over the considered time period. (*B*): distributions for sexual contacts network.

2.3 Correlation between loyalty and degree

Degree and loyalty, while not being independent variables, are nonetheless not trivially correlated. Fig. S3 shows the scatter plots between the degree of a node in configuration c and its loyalty for the pair of configurations $c, c + 1$, for both networks. For each value of k, θ is found to range over a wide interval. This is clearly visible up to $k \approx 10^2$ for the cattle trade network, and $k \approx 10$ for the sexual contact network. Higher degree nodes are much less frequent, so the statistics becomes poorer and the heterogeneity in θ decreases as k increases. Pearson correlation coefficients are found to be low for both networks (0.04 for the cattle trade network and 0.15 for the sexual contact network), consistently with the observed large variations. They are however significantly larger than the coefficients of the null model: 95% confidence interval of $(-0.002, 0.002)$ and $(-0.006, 0.007)$, for the cattle trade network and the sexual contact network, respectively. This points to a positive, albeit weak, correlation between degree and loyalty. The confidence intervals for the null model are obtained by randomly shuffling several times the sequence of θ's, in order to highlight any spurious correlation with the degree sequence.

3 Loyalty and other similarity measures

We analyze here the relationship between loyalty and other possible measures of similarity of the neighbor structure of a node across time. Firstly we consider a measure introduced as *social strategy* in [1]. In our context, if we call $\tilde{k}_i^{1,c}$ the (in-)degree of node i in the network resulting from the aggregation of snapshots 1 to c, then $i's$ social strategy in those configurations will be computed as $\gamma_i^{1,c} = \tilde{k}_i^{1,c} / \left(\sum_{c'} k_i^{c'} \right)$. $k_i^{c'}$ is as usual the (in-)degree of i in configuration c'. This definition is the same as in [1], except for a normalizing factor c. We make this choice in order to make the comparison with θ more straightforward. The most important qualitative
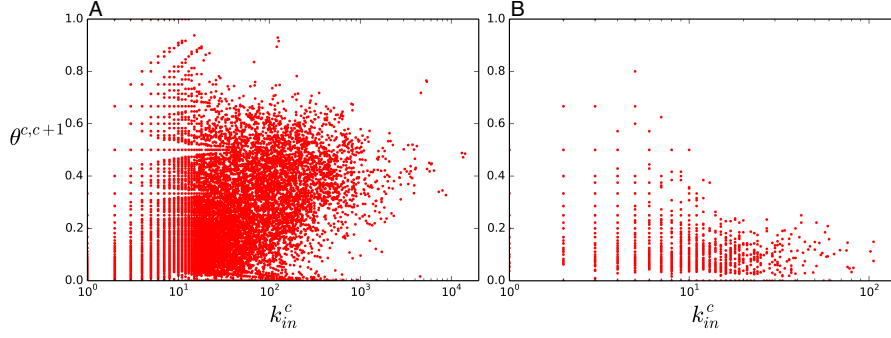
3

Figure S3: Scatter plots showing degree at configuration c vs loyalty between configurations $c, c+1$. Each point represents a node. (*A*): cattle network. (*B*): sexual contacts.

difference between loyalty and social strategy is that the former is always computed between a pair consecutive snapshots, while the latter typically describes an average behavior computed on several configurations (from 1 to c in our notation). Indeed only in the trivial case of γ computed on just two snapshots, loyalty and strategy are univocally related: $\gamma_i^{1,2} = \frac{1}{1+\theta_{1,2}^i}$. In general, $\gamma_i^{1,c}$ will be a non trivial combination of all the consecutive loyalties $\theta_i^{1,2}, \theta_i^{2,3} \cdots \theta_i^{c-1,c}$ and degrees. Fig. S4A shows the correlation between social strategy in cattle network, computed from 2006 to 2010, and loyalty between $2009, 2010$.

We now consider a measure of neighbor similarity derived from Pearson correlation coefficient. This measure is analogous to what is called *adjacency correlation* in [2]. For each node we build two vectors, v_i^c, v_i^{c+1}, of dimension $|\mathcal{V}_i^c \cup \mathcal{V}_i^{c+1}|$, i.e. these vectors will contain an entry for each vector that is neighbor of i in at least one of the two configurations. v_i^c has entries equal to 1 for nodes that are in $\mathcal{V}_i^c$, and zero otherwise, and the same for v_i^{c+1}. We than consider the Pearson correlation coefficient between the two vectors, $\xi^{c,c+1}$. This can be directly related to the loyalty $\theta_i^{c,c+1}$ and the degrees of the node in the two configuration k_i^c and k_i^{c+1} through the formula

$$\xi^{c,c+1} = -\frac{k^c + k^{c+1}}{\sqrt{k^c k^{c+1}}} \frac{1}{1+\theta^{c,c+1}} \sqrt{(1+\theta^{c,c+1})^2 \frac{k^c k^{c+1}}{(k^c + k^{c+1})^2} - \theta^{c,c+1}} \qquad (5)$$

In the above equation we have omitted the subscript i: $\xi^{c,c+1} = \rho_i^{c,c+1}$, $k^c = k_i^c$ and $\theta^{c,c+1} = \theta_i^{c,c+1}$. Fig. S4B shows the scatter plot $\xi^{c,c+1}$ versus $\theta^{c,c+1}$. We see that, due to the definition of vectors v^c, $\xi \in [-1, 0]$. This formula can be simplified if we need just an average behavior: assuming $k^c = k^{c+1} = k$, where k is the average connectivity, the formula reduces to $\langle \xi^{c,c+1} \rangle = -(1 - \theta^{c,c+1})/(1 + \theta^{c,c+1})$. From this we get that $\theta = 0$ (no memory) corresponds to $\xi = -1$, while $\theta = 1$ (perfect memory) corresponds to $\xi = 0$.

Finally, we analyze an application of cosine similarity. For each node vectors v_i^c, v_i^{c+1} are built as before. Then cosine similarity between those vectors is defined as $\zeta = v_i^c \cdot v_i^{c+1} / (|v_i^c||v_i^{c+1}|)$. It can be shown that, like ξ, ζ can be written in terms

4

of degree and loyalty:

$$\zeta^{c,c+1} = \frac{\theta}{1+\theta} \frac{k^c + k^{c+1}}{\sqrt{k^c k^{c+1}}} \tag{6}$$

The average behavior this time is $\langle \zeta^{c,c+1} \rangle = 2\theta^{c,c+1}/(1 + \theta^{c,c+1})$ (see scatter plot in Fig. S4C).

In conclusion, social strategy, being computed on a sequence of more than two configurations, represents a qualitatively different measure with respect to loyalty, albeit the two measures being correlated (see Fig. S4A). On the other hand, both Pearson ξ and cosine similarity ζ can be completely determined in terms of degree and loyalty. Moreover, the mean trend is well modeled by the averaged version of these measure, which discounts degree (see Fig. S4).
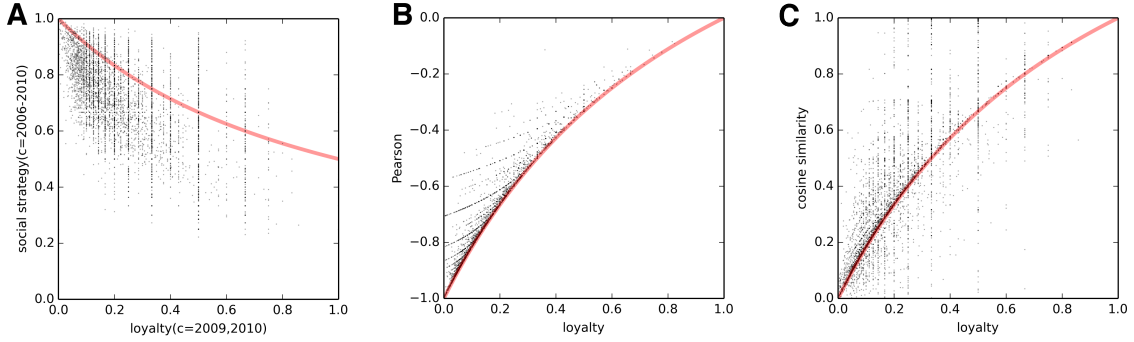

Figure S4: Cattle network: correlation between loyalty and other neighbor similarity measures. *(A)*: scatter plot showing social strategy (γ) computed from 2006 to 2010 vs loyalty between 2009, 2010. Each point represents a node. The red line represents $\gamma_i^{1,2} = \frac{1}{1+\theta_{1,2}^i}$; Pearson correlation is -0.59. *(B)*: Pearson (ξ) vs loyalty. The red line represents $\langle \xi^{c,c+1} \rangle$. *(C)*: cosine similarity ζ vs loyalty. The red line represents $\langle \zeta^{c,c+1} \rangle$.

4 Modeling infection potentials

Infection potentials π_D and π_L are modeled with a sum of an exponential distribution, to account for the behavior at $\pi \simeq 0$, and a Landau distribution, to mimic the particular asymmetry around the peak. The exact formulation is the following:

$$f(x; \mu, \sigma, r, q) \propto \exp(-qx) + r \int_0^\infty dt \, \sin(2t) \exp\left[-t\frac{x-\mu}{\sigma} - \frac{2}{\pi} t \log t \right]. \tag{7}$$

There are four free parameters: one for the exponential distribution, two for the Landau distribution, and one driving the relative importance of one function with respect to the other. An overall scaling coefficient is fixed by normalization.

5 Robustness of the risk assessment procedure in varying parameters and assumptions

5.1 Threshold ϵ

In the following we examine the behavior of the infection potentials π_D and π_L in varying the value of the threshold. Fig. S5 shows that in the cattle trade network the peak position of π_D increases with ϵ, from 0.3 to 0.6. Such behavior is present in the sexual contact network too, albeit less evident (from 0.3 to 0.5). Unlike π_D, π_L distributions remain stable as ϵ varies. As a result, the probability of a loyal node being infected (π_L) does not depend on the choice of ϵ. The choice of threshold $\epsilon = 0.1$ thus allows to maximize the distance between π_D and π_L distribution while preserving enough statistics for the loyal nodes.
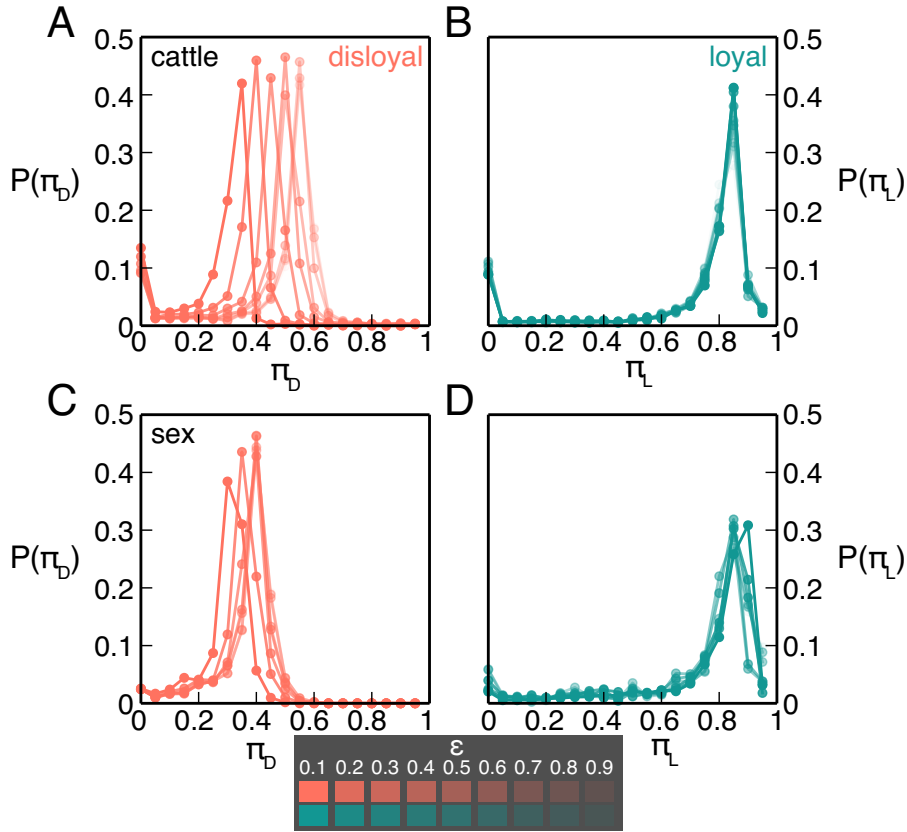
Figure S5: Behavior of infection potentials π_D and π_L as ϵ varies. (*A*),(*C*): π_D curves. (*B*),(*D*): π_L curves. (*A*),(*B*): cattle network. (*C*),(*D*): sexual contacts.

It is important to note that the value of ϵ also affects the transition probabilities T_{DD}, T_{LL} in their functional dependence on the degree (Figure 3C,D of the main text). For each threshold value, such dependence needs therefore to be assessed through a fitting, to be used for the prediction of the loyalty values in the unknown network configuration.

Figure S6: Risk ratio (ν) distribution for cattle network. (A): $\mathcal{I}_{s,h}^c$ as set of top 50% highest ranking nodes. (B): $\mathcal{I}_{s,h}^c$ as set of nodes with $\rho > 0.6$.

5.2 Definitions for the risk ratio ν

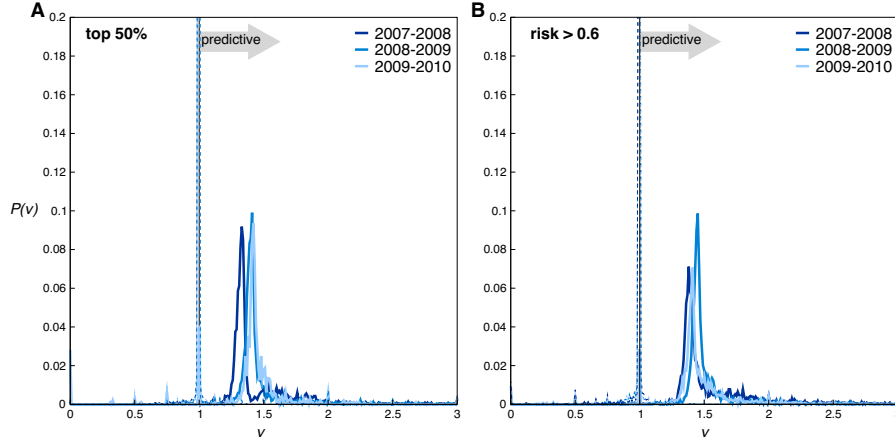In the main paper the risk ratio ν is computed considering the set $\mathcal{I}_{s,h}^c$ of the top 25% highest ranking nodes. Here we explore two different ways of defining this quantity:

- $\mathcal{I}_{s,h}^c$ as the set of the top 50% highest ranking nodes (Fig. S6A);

- $\mathcal{I}_{s,h}^c$ as the set of nodes with epidemic risk $\rho > 0.6$ (Fig. S6B).

Results are reported in Fig. S6 showing the invariance of the observed ν results on this arbitrary choice.

5.3 Definition of the early stage of an epidemic

In the main paper we consider an initial stage of the epidemic up to $\tau = 6$. This choice being arbitrary, it is informed by the simulated time behavior of the incidence curves (see Fig. S7) and the aim to focus on the initial stage of the epidemic.

We also tested a longer initial stage ($\tau = 10$) for the sexual contacts network, to assess the impact of this variation on the obtained results. We obtain distributions of the infection potential, of the relative risk ratio, and of the predictive power showing sharper peaks, however with unchanged peak positions (Fig. S8 for the sexual contact network). Peaks are expected to be sharper, because with $\tau = 10$ a larger fraction of the network is reached by the outbreak. The fact that peak positions do not change, however, reveals that we are able to provide accurate epidemic risks already at the earlier phase of the epidemic ($\tau = 6$), when such information is mostly needed.

5.4 Aggregation time window

The choice of yearly aggregation time in the case of the cattle trade network is informed by its annual seasonal dynamics; the six-months aggregating window for the sexual contact network is instead arbitrary. Here we explore other aggregating
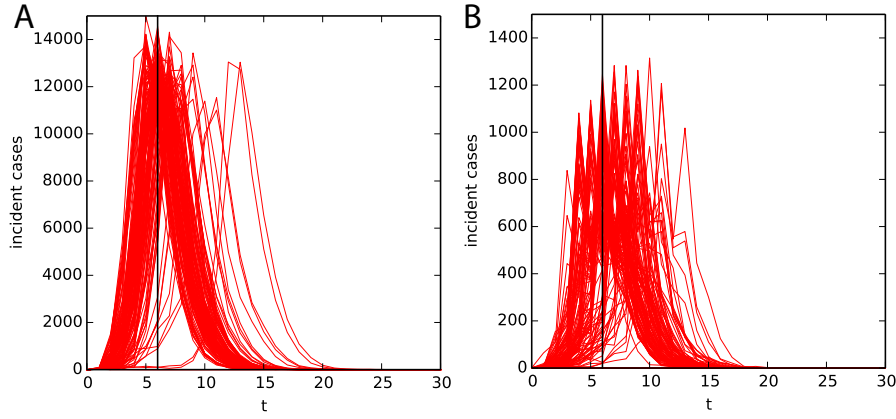
Figure S7: Simulated incidence curves obtained by changing seeding node and network configuration for the cattle trade network (*A*), and the sexual contacts network (*B*). Black line indicates $\tau = 6$.
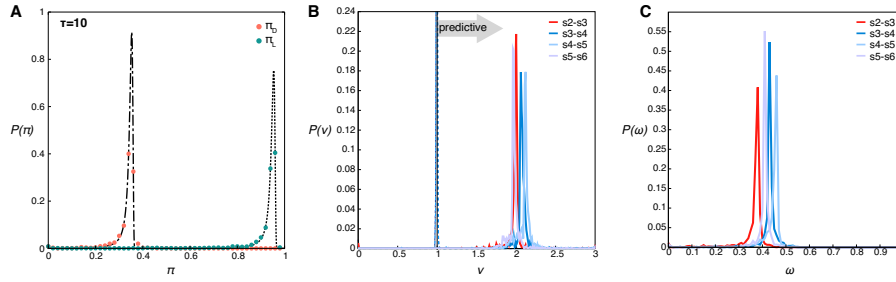
Figure S8: Invasion stage of the outbreak up to $\tau = 10$ for sexual contacts network: distribution of the infection potentials (*A*), the risk ratio ν (*B*) and the predictive power ω (*C*).

windows for both networks to explore the impact they may have on the obtained results.

We consider configurations for the sexual contact network consisting of 3-months aggregation. When calculating the risk ratio and the predictive power (Fig. S9B,D), we find distributions similar to the ones reported in the main text, with unchanged peak positions. The distributions however appear to be noisier, especially as far as ω is concerned, likely induced by the increased sparseness of the network configurations.

We also try a different aggregation time for cattle network: 4-month windows. Risk ratio and predictive power distributions are presented in Fig. S9A,C. We observe that ω is on average quite low: this is likely due to the fact that aggregation windows shorter than one year fail to take into account the seasonal patterns, thus decreasing system memory.

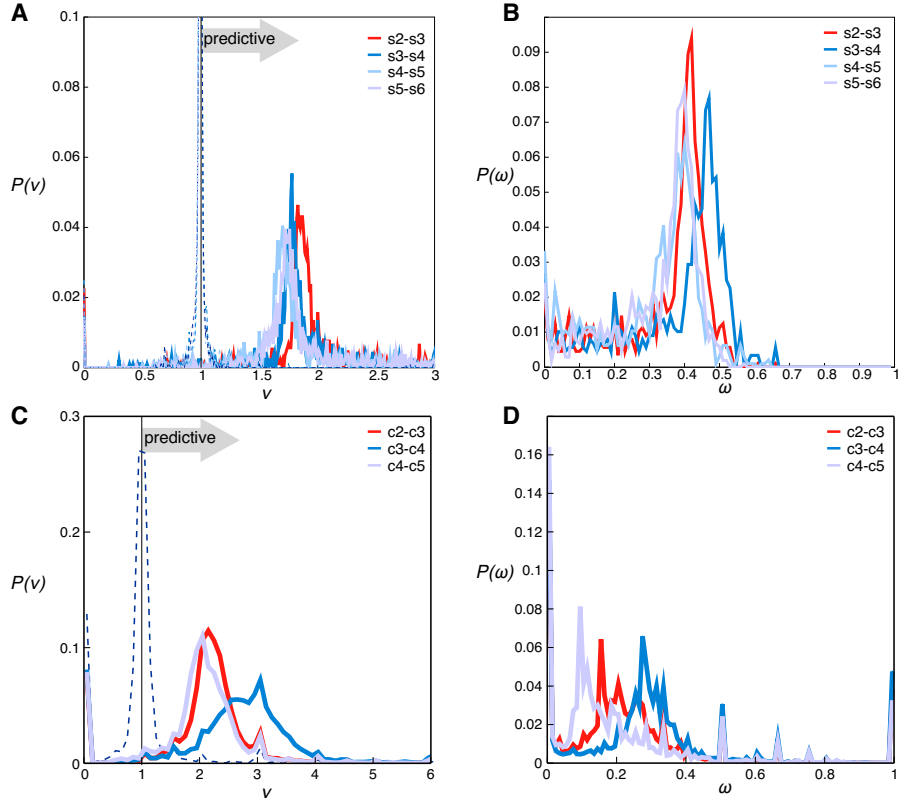Figure S9: Exploration of different aggregating windows (cattle: 4-month, sex: 3-month). Distributions of risk ratio ν in cattle (A) and sexual contacts (C). Distribution of predictive power ω in cattle (B) and sexual contacts (D).

6 Memory driven model: analytical understandings

6.1 Amount of memory

In the following we analytically quantify the amount of memory in the memory driven model as the probability $f_{c,c+1}$ that a link present in configuration c is also present in configuration $c+1$. This can be expressed as:

$$f_{c,c+1} = (1-d)\left[p_\alpha + \frac{1}{N}\frac{b(1-d)}{b+d}\frac{\zeta(\gamma-1)}{\zeta(\gamma)}\right], \tag{8}$$

where the first term, $(1-d)p_\alpha$, is the probability of remaining active and at the same time keeping a particular neighbor. The second term is the probability of not keeping a neighbor but recovering it with one of the new stubs. ζ is the Riemann ζ-function. $f_{c,c+1}$ can indeed be interpreted as the system memory, as it is a good estimator of the fraction of links that survive from one configuration to the following.

The second term in Eq. 8 is suppressed by $1/N$ and can be disregarded in our case given the large size of the networks ($N = 10^4$). $f_{c,c+1} \approx (1-d)p_\alpha$ therefore provides a first order approximation that correctly matches the numerical results (see Fig. S10A for the comparison).

6.2 Probability associated to zero loyalty

The probability of a node with in-degree h_c having zero loyalty ($\theta_{c,c+1} = 0$) can be computed analytically as

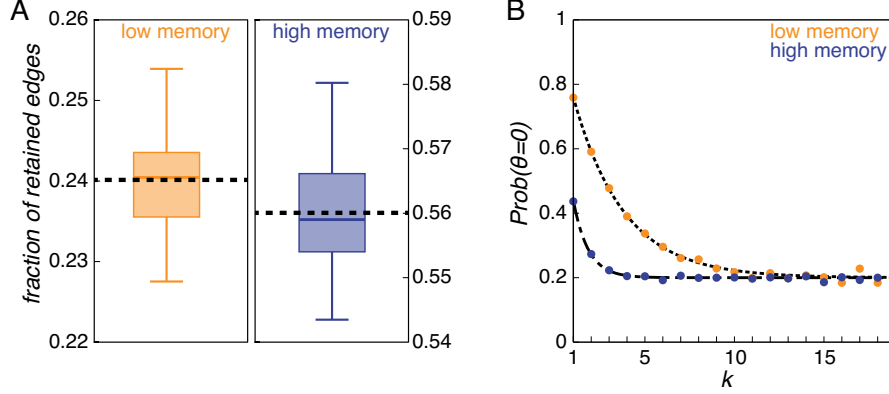$$P(\theta_{c,c+1} = 0|k_c) = d + (1 - d)(1 - p_\alpha)^{k_c}. \tag{9}$$

Figure S10: Characterization of the memory driven dynamical model. (*A*): the memory of the system, in terms of the fraction of edges retained from one configuration to the following. Boxplots represent median and quartile positions. The distributions are computed over 50 realizations of the model. Dashed lines represent the theoretical prediction. $p_\alpha = 0.3, 0.7$ for low and high memory, respectively. (*B*): probability for a node with a given in-degree k to be completely disloyal ($\theta = 0$) between two following snapshots. Points represent numerical simulations, while lines show the theoretical estimates.

In Fig. S10B we check this result against numerical simulations.

7 Memory driven model: additional properties

In the main paper the transitions probabilities between loyalty statuses are shown only for the real networks (main paper Fig. 3C and 3D). Here we present them for the memory driven model. Fig. S11 reports these probabilities in case of low and high memory, along with the modeling functions.

In addition, we explore different values of the model parameters and discuss the changes in the network properties. In particular, we explore different values for the probability of becoming active (b) or inactive (d), other than the choice used in main paper ($b = 0.7$, $d = 0.2$). Fig. S12A, S12B, S12C are the equivalent of main paper Fig. 5A, and show the in-degree distribution for different values of b, d in the set $\{0.2, 0.7\}$. $P(k_{in})$ is very robust when changing these parameters, and in all cases follows the slope of the β_{in} distribution. Fig. S12D, S12E, S12F are the equivalent of main paper Fig. 5B, and show the loyalty distributions. We observe that the overall shape is insensitive to parameters change. There is however, a tendency to
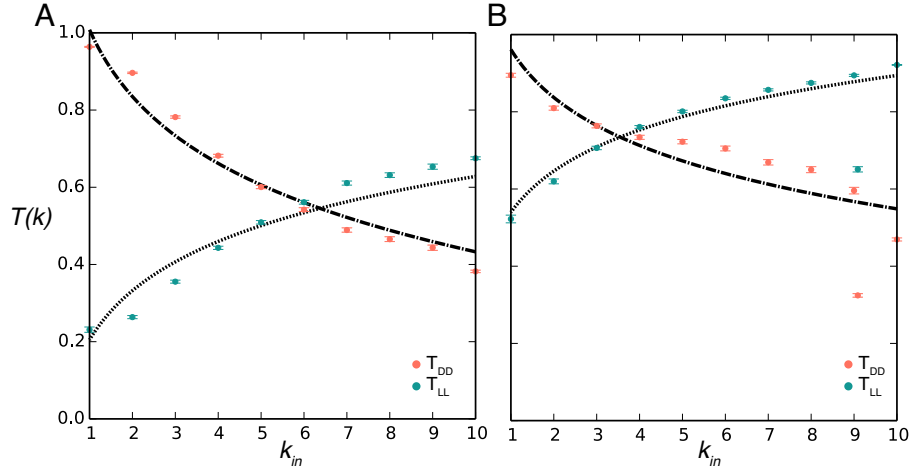
Figure S11: Memory driven model: loyalty transition probabilities between loyal statuses ($T_{LL}(k)$, green) and disloyal statuses ($T_{DD}(k)$, orange) as functions of the degree k_{in} of the node. (*A*): low memory model ($p_\alpha = 0.3$), (*B*): high memory model ($p_\alpha = 0.7$). Dashed lines represent the logarithmic models: $T_{DD}(k) = 1.01 - 0.25 \log k$, and $T_{LL}(k) = 0.20 + 0.18 \log k$ for the low memory; $T_{DD}(k) = 0.96 - 0.17 \log k$, and $T_{LL}(k) = 0.53 + 0.15 \log k$ for high memory. Error bars represent the deviation $\pm \{T(k)\left[1 - T(k)\right]/N_k\}^{1/2}$, where N_k is the number of nodes with degree k used to compute $T(k)$. Last value for k: $k = 10$ includes all nodes with degree equal or higher.

have higher θ values for low b, d. This is to be expected, since higher probabilities of going from active to inactive and vice versa mean larger turnover, which leads to lower memory and therefore lower overall loyalty.

8 Validation in the stochastic case

We repeat the analysis reported in the main text by considering a stochastic Susceptible-Infectious approach. Given the same initial conditions, we perform r different stochastic runs, each leading to potentially different outcomes. For each node i, we compute the fraction $f_i(s)$ of runs that node i is infected from epidemics starting from seed s within time step τ. For validation, we need to compare the list $\{\rho_i\}(s)$ of the node epidemic risks computed with our methodology with the list $\{f_i\}(s)$ of the probabilities of actually getting infected. If our estimated risks are reliable, then the two lists need to be correlated, as a higher risk should correspond to a higher probability to get infected. In order to evaluate this, we compute the Pearson correlation coefficients between $\{\rho_i\}(s)$ and $\{f_i\}(s)$, for each possible seed s. The list of these coefficients can then be summarized in a distribution. Fig. S13B and S13D show such distributions for the sexual contact network for two different values of the infection transmissibility (0.75 and 0.85, respectively). In order to check that the correlation coefficients are significantly different from zero, we compute the same distributions after reshuffling the epidemic risks (dashed lines in plots). Fig. S13A
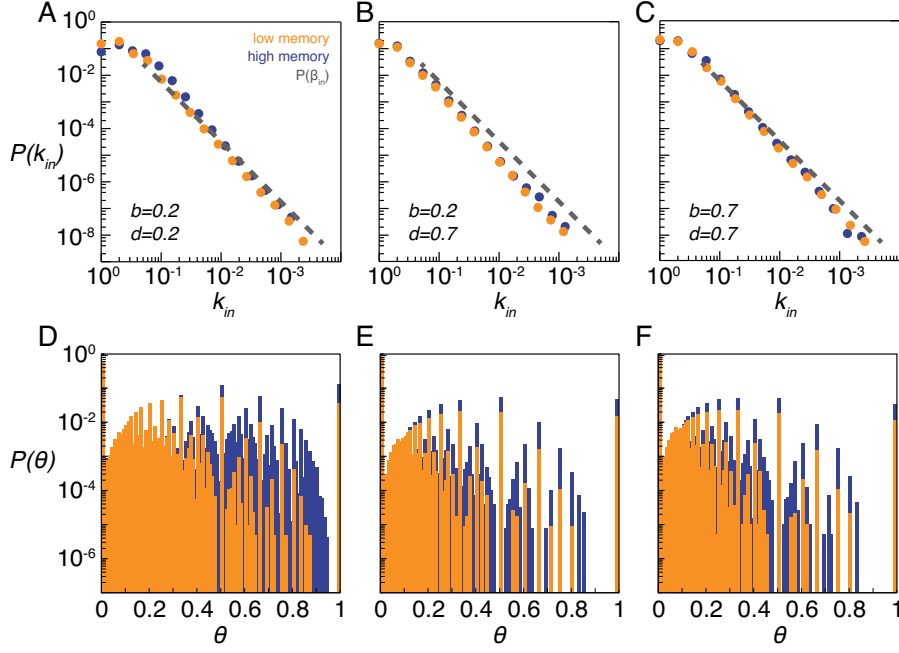
Figure S12: Memory driven model: degree and loyalty distribution when changing the probabilities of becoming active or inactive. (A),(B),(C): in-degree distributions when $(b, d) = (0.2, 0.2), (0.2, 0.7), (0.7, 0.7)$, respectively. (C),(D),(E): loyalty distributions for the same parameter configurations.

and S13C are the equivalent of Fig. 3B in main paper and show that the peak position of the infection potential does not change from the deterministic case. Noise and peak width, however, increase considerably, as well as the probability of having $\pi_D = 0$, and this effect is more pronounced for lower infection transmissibilities.

9 Cattle network: taking into account links weights

Links in cattle network can be assigned a weight attribute in terms of the number of moved animals. These additional data can be included in the modeling of diseases spread, assuming that larger batches have a greater probability of carrying the disease from the source holding, to the destination. This feature is included in the disease model, by assuming a per-animal transmissibility λ. Then, given a movement of w animals, the transmission probability along that link will be $[1 - (1 - \lambda)^w]$ (same approach as in SI of [3]). Loyalty needs to be generalized to the case of weighted network, too. The most straightforward generalization is obtained by considering the quantities in Eq. (2) of main paper $\mathcal{V}_i^{c-1}, \mathcal{V}_i^c$ as multisets (see, for instance, [4]), where each neighbor appears as many times as the weight of the corresponding link. Then the *weighted loyalty* on the weighted network is defined, as before, by Eq. (2) of main paper, using the definitions of multiset union and intersection: $\mathcal{V}_i^{c-1} \cup \mathcal{V}_i^c = \sum_j \max\left(w_{ji}^{c-1}, w_{ji}^c\right)$ and $\mathcal{V}_i^{c-1} \cap \mathcal{V}_i^c = \sum_j \min\left(w_{ji}^{c-1}, w_{ji}^c\right)$, where w_{ji}^c is the weight of the link $j - i$ in configuration c (assuming $w = 0$ if no such link is present). Other choices of similarity between sets of neighbors are
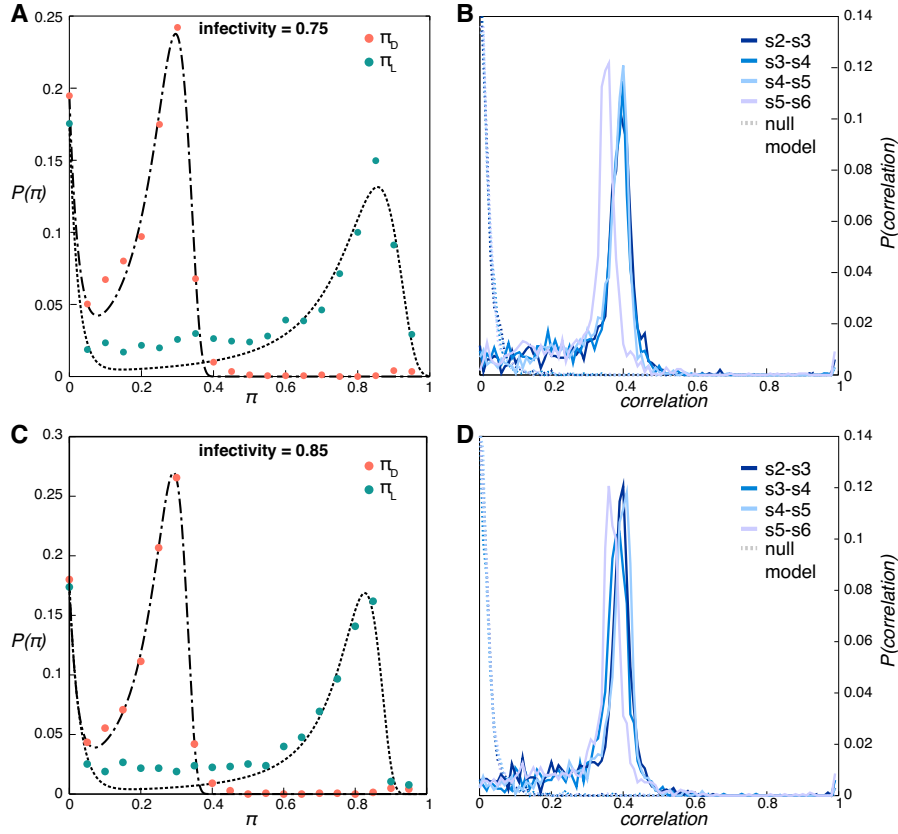
Figure S13: Applying the methodology to sexual contact networks using a stochastic epidemic model. $(A),(C)$: infection potentials for infectivity 0.75 and 0.85, respectively. $(B),(D)$: distribution of the Pearson correlation coefficient between the computed epidemic risks and the probability of actually being infected, for infectivity 0.75 and 0.85, respectively. Dashed lines show distributions from the null model.

possible, however this one is the most natural generalization, since it has a very similar distribution to the unweighted loyalty (Fig. S14A), and correlates well with it (Fig. S14B). We now compute the infection potentials and then the epidemic risks, using this new loyalty. We validate the computed risks analogously to what we did in Sec. 8. Results are presented in Fig. S15, showing the generalizability of our approach to the weighted case too.

10 Assessing the robustness of risk based prediction with respect to simple predictors

We have shown that ρ effectively represents the risk of being infected, as shown in the Validation section of main paper. We now show that ρ is a significant improvement in prediction accuracy, with respect to simpler measures, like the degree of a node. From configurations $c-1, c$ of cattle network we compute the risk of being

Figure S14: Weighted cattle network: extending the definition of loyalty. *A* shows the cumulative distributions for the unweigthed loyalty (green) and multiset loyalty (red). Different tones of colors refer to different network c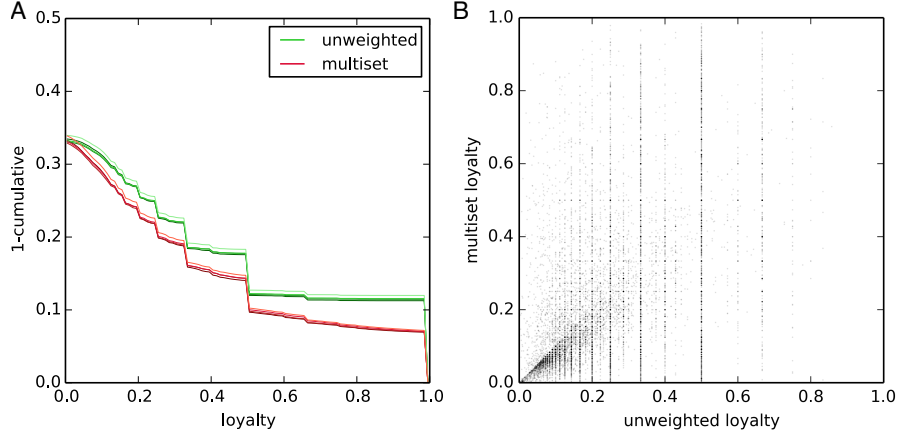onfigurations. *B* Scatter plot correlating the unweighted loyalty and the multiset loyalty. Pearson correlation coefficient is 0.92.

infected at $c+1$: $\rho_i = \rho_i^{c+1}(s)$, as in Eq. (3) of main paper. For each node i for which we can compute ρ_i we then have the binary variable *outcome* indicating if node i is eventually hit by the epidemic in configuration $c + 1$. We perform a multivariable logistic regression to check that ρ is actually a predictor for *outcome*, adjusting the in-degree in configuration c: k_i^c. In particular, due to the high heterogeneity of k, we adjust for the log of the degree. Tab. S1 shows the results of the performed regressions. As the crude odds ratios show, both ρ and k^c, on their own, are meaningful predictor of infection in configuration $c + 1$. We are however interested in assessing wether our risk is still a predictor, once the effect of knowing the degree is discounted for. The odds ratio for ρ adjusted for degree is still significantly greater than one, meaning that even within nodes of the same degree, nodes at high risk are likelier to get infected. In other words, computing the risk (for which the knowledge of the degree of the node is needed) gives more predicting power than the sole knowledge of degree.

	crude OR	adjusted OR
log(degree)	2.88 $[2.87, 2.89]$	2.08 $[2.07, 2.10]$
risk	4.82 $[4.78, 4.86]$	2.50 $[2.49, 2.51]$

Table S1: Odds ratios of being infected in configuration $c + 1$, given degrees in c and computed risks. Crude odds ratios refer to two separate univariate regressions; adjusted odds ratios are obtained through a single multivariate regression. 95% confidence intervals are reported.
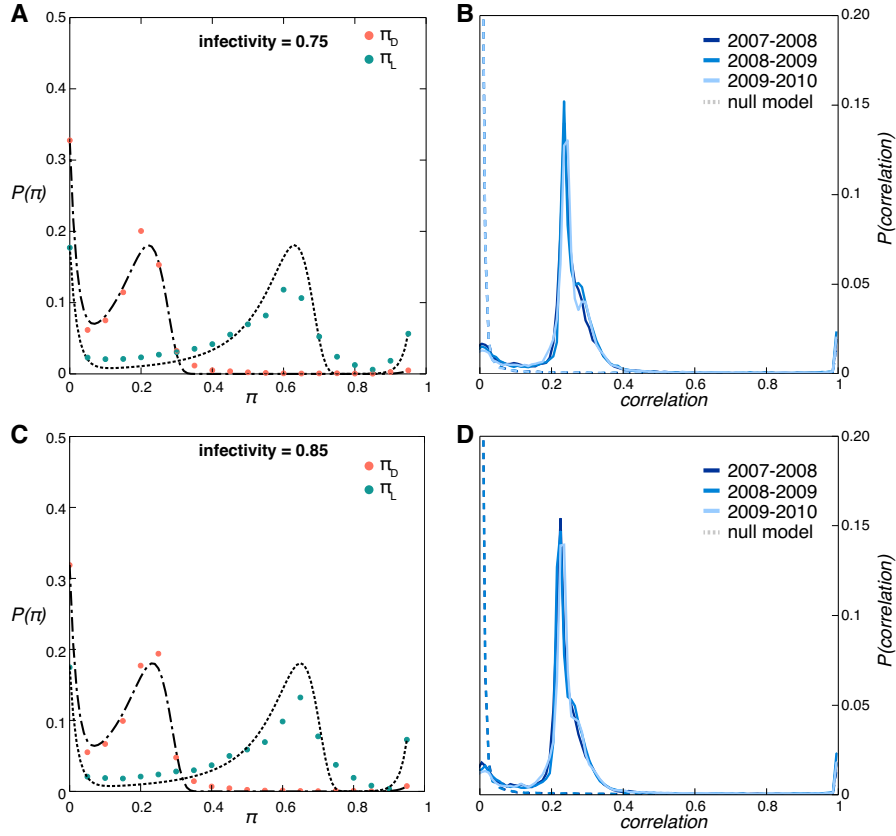
Figure S15: Weighted cattle network: risk prediction computation and validation. (A),(C): infection potentials for single-animal infectivity 0.75 and 0.85, respectively. (B),(D): distribution of the Pearson correlation coefficient between the computed epidemic risks and the probability of actually being infected, for infectivity 0.75 and 0.85, respectively. Dashed lines show distributions from the null model.

11 Application to human proximity networks

The main difficulty in applying our methodology to physical proximity networks in human is that generally those networks are much smaller than the ones we have examined, that making it difficult to reach enough statistics to fit the form of infection potentials and transitions probability, and then perform the validation. We show here how we can overcome these impairments and apply successfully our strategy to a network of face-to-face proximity at a scientific conference, collected by the Sociopatterns group [5]. This network records the interactions of 113 nodes during a period of 2.5 days. We split such networks in 30 configurations (corresponding to hourly time steps), and use the first 29 configurations to train our methodology, in order to give predictions on the $30th$. We use this large number of configurations in order to be able to build reliable empirical distributions for the infection potentials and the transition probabilities between loyalty statuses. Once risks are computed as usual, it is not possible, however, to perform the validation as we did for cattle, sexual contacts and memory driven models. This impossibility arises from the

fact that the computed risk ratios are too few to build their distribution. In order to validate our methodology we therefore use the same technique implemented in Sec/ 10: for every node, we compute the odds ratio of being infected in the last configuration, given the knowledge of degree and the computed risk. Results are reported in Tab. S2. Computer risks are strong predictors for infection, even after adjusting for degree. Moreover, unlike cattle network (see Tab. S1), degree alone is not a predictor. Predictive power ω is on average high: median 0.87, with quartiles $Q_1 : 0.69, Q_2 : 0.97$.

	crude OR	adjusted OR
log(degree)	1.16 [1.13, 1.20]	0.95 [0.89, 1.02]
risk	11.97 [7.79, 18.4]	22.34 [7.90, 63.3]

Table S2: Odds ratios of being infected in last configuration last, degree and computed risk. Crude odds ratios refer to two separate univariate regressions; adjusted odds ratios are obtained through a single multivariate regression. 95% confidence intervals are reported.

References

[1] Miritello G, Lara R, Cebrian M, Moro E (2013) Limited communication capacity unveils strategies for human interaction. Sci Rep 3, 1950.

[2] Clauset A, Eagle N (2012) Persistence and periodicity in a dynamic proximity network, ArXiv:1211.7343.

[3] Bajardi P, Barrat A, Savini L, Colizza V (2012) Optimizing surveillance for livestock disease spreading through animal movements. J Roy Soc Int (June, 2012).

[4] Stanley R P (1997). Enumerative Combinatorics, Vols. 1 and 2., Cambridge University Press.

[5] Isella L, Stehlé J, Barrat A, Cattuto C, Pinton J-F, Van den Broek Wouter (2011) What's in a crowd? Analysis of face-to-face behavioral networks, Journal of Theoretical Biology 271,1:166-180.

END

Conclusions and perspectives

Time evolution of contacts within a population has a notable impact on the conditions that drive the spread of a disease. In this thesis we have developed new tools to deal with the interplay between contact and disease dynamics, in order to provide vulnerability assessment strategies that can be applied to real settings and realistic disease models.

Using a multilayer approach, we have introduced an analytical methodology to compute the epidemic threshold on a generic temporal network, that turns into a tool for predicting the outcome of pathogen introduction in a wide variety of contexts. We have developed this computation in order to include different disease features, like latency and immunity, and different representations of temporal networks, both in discrete and continuous time. In addition, we have gauged the impact that two important features of empirical temporal networks – time resolution and data collection interval – have on the estimation of the threshold.

We have introduced several datasets that track contacts in various contexts, all epidemiologically relevant. These range from face-to-face human interactions, useful to study the spread of airborne diseases, to sexual contacts, along which sexually transmitted infections can naturally spread. In addition, we have presented an analysis of the network of cattle trade movements in three European countries – Hungary, Italy and Sweden –, which is characterized by several temporal and spatial patterns, all relevant for the spread of diseases.

Finally, we have concentrated our analysis on specific real settings and classes of diseases. We have used the computation of the threshold to gauge the vulnerability of the Italian cattle trade network, in both space and time. We have then turned to human infectious diseases spread via sexual contacts, and assessed under which circumstances latency period has a positive (or negative) impact on the vulnerability of the network, with a particular focus on syphilis.

We then remarked that epidemic threshold provides a measure of global vulnerability of the network, while it cannot provide targeted risk predictions for specific hosts. In

order to overcome that, we have developed a data-driven methodology for predicting individual infection risk in a temporal network. In addition, given that in many real settings we have only data describing past contact configurations, we have devised new techniques that allow us to compute such risk for future potential outbreaks, by relying only on past network data. We have assessed the performing power and generality of this methodology, by testing it on both sexual contacts and cattle network. We have then devised a generative network model to test which dynamic features our methodology is sensitive to.

From the theoretical point of view, the framework we have developed is general enough to deal with generic networks and various disease characteristics. In the future, however, it should be further extended to include other features found in the progression patterns of many diseases, like heterogeneous infectious periods among different hosts. In addition, in order to get even more accurate numerical estimates of the threshold, it will be necessary to go beyond the mean field assumptions that underlie the computation of the threshold both in the static and the temporal case. Finally, the epidemic threshold provides the condition that discriminates between epidemic outbreak and disease extinction, but it provides no information on other important measures like outbreak size, attack rate, and endemic prevalence. Further research is needed to extract the value of these quantities from the equations describing the epidemic process. This is however impaired by the highly non-linear nature of those equations, and the development of new modeling and mathematical tools will be needed.

The issue of predicting targeted epidemic risk from past data is extremely relevant in many contexts. In our work we have provided a new technique to achieve that in various contexts. We however assume simple disease models, that will have to be made more realistic, in order to be applied to specific epidemics. In addition, while our methodology is very general, it could be made more accurate in some context if it were able to include further data which are indeed context-specific, like geographic meta-data in cattle trade networks.

A. Appendix

A.1. Networks - essential glossary

graph

A graph is an entity composed of a set of elements called nodes (or vertices), and a set of pairwise relationships among nodes, called links (or edges). Links can be symmetric relations (undirected graph), or directed (directed graph).

adjacency matrix

Let us assume a graph of N nodes: $i = 1 \cdot N$. We define a matrix A such that $A_{ij} = 1$ if a link exists from i to j. We can map each node i of the graph onto e_i, the canonical basis vector of $\mathbb{R}^N$. Then A naturally behaves as a linear map $\mathbb{R}^N, \mathbb{R}^N \to \mathbb{R}$. If the graph is undirected, then A is symmetric. If A takes non-zero values other then 1, then the corresponding graph is weighted, and link $i - -j$ has weight A_{ij}. Usually weights are indicated with letter w. A weighted adjacency matrix is sometimes called W.

degree and strength

Let us take a graph with N nodes. Degree (usually called k) of node i is the number of neighbors of node i, i.e., the number of links with one tip in i. If the network is weighted, in addition to degree we define strength (s) as the sum of the weights of these links. If the graph is directed, we define k^{in} as the number of incoming connections, and k^{out} as the number of outgoing connections. The same goes for strength. As generally as possible (directed, weighted) in terms of adjacency matrix, $s_i^{in} = \sum_j A_{ji}$; $s_i^{out} = \sum_j A_{ij}$. For an unweighted, undirected graph, we just have $k = \sum_j A_{ij}$.

distributions

Given a graph of N nodes, we have a degree list $\{k_i\}$. Out of that, we can build a histogram of a certain degree occurring, i.e., a degree distribution. Conversely, given a fixed distribution, we can build an ensemble of graphs whose degree sequences are samples of that distribution, through the *configuration model* [219]. The same holds for strengths, and directed graphs.

paths and reachability

A path of length l in a graph is an ordered subset of l edges such that each edge ends in the node where the following edge starts from. The shortest path between two nodes is the shortest among the paths that start from one node and end in the other. A node i is reachable from j if there exists at least one path going from j to i.

connected components

A connected component of a graph is a subset of nodes that are all reachable one from the other. If this is true in a directed network, then this said is called a strongly connected component (SCC). A weakly connected component of a directed graph is instead a subset that is a connected component after the graph is symmetryzed, i.e., all edges are made undirected.

clique

A clique is a fully connected subgraph, i.e., a subset of nodes in which every one is connected to every one else. Commonly, when we talk about cliques as well as connected components, we assume they are maximal, namely adding a further node to the set would break its property.

A.2. Tensor representation of temporal networks

The tensor formulation of multilayer networks has been put forward in [102]. Here we restrict it to temporal networks. Let us assume a network of N nodes. We start by assuming it doesn't evolve in time. Let V be the vector space spanned by the basis $\{e_1, e_2, \cdots, e_N\}$ which is in one-to-one correspondence with the nodes in the network (see Appendix A.1). In particular, $V \simeq \mathbb{R}^N$. We consider the dual space V^* as the

vector space spanned by the linear maps $\{e_1^*, e_2^*, \cdots, e_N^*\}$. They are defined as follows: $V \to \mathbb{R}$ and $e_i^*(e_j) = \delta_{ij}$. This is a traditional definition of dual space of a vector space. Then, we can define the adjacency matrix A as a multilinear map (i.e., a tensor) $A : V^* \otimes V \to \mathbb{R}$, defined as $A = \sum_i \sum_j A_{ij} e_i^* \otimes e_j$, with component $A_{ij} = A(e_i, e_j^*)$ encoding the value of link $i - j$. A thus becomes a rank 2, 1-covariant, 1-contravariant tensor. The generalization to the temporal case is now straightforward. Let us assume we have T snapshots. We then consider the space $W \simeq \mathbb{R}^T$ spanned by the basis $\{f_1, f_2, \cdots, f_T\}$ which is in one-to-one correspondence with the snapshots. Then we define the adjacency tensor as a rank 4 tensor $A = \sum_t \sum_s \sum_i \sum_j A_{ts,ij} f_t^* \otimes f_s \otimes e_i^* \otimes e_j$, with component $A_{ts,ij} = A(f_t, f_s^*, e_i^*, e_j)$ encoding the value of link from node i at time t, to node j at time s. This allows for both intra-layer (same time) and inter-layer (different times) links, as shown in Fig. 2.2.

A.3. Discrete Fourier Transform

We start by considering Fourier transform of a function $f \in L^2(\mathbb{R})$, where $L^2(\mathbb{R})$ is the set of $\mathbb{R}$-valued functions for which $\int_{-\infty}^{\infty} dx \, |f(x)|^2 < \infty$:

$$F(k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx f(x) e^{-i2\pi kx}. \tag{A.1}$$

$F(k)$ is the component of f along the basis vector $e^{i2\pi kx}$ of $L^2(\mathbb{R})$, i.e., it tells us what is the contribution to f of the harmonic k. When we have an activity timeline, instead of having a continuous function f, we have a list of values x_j, with $j = 0, \cdots, n-1$. Hence, now harmonics are discrete and finite: $k = 0, \cdots, n-1$. They can be compute through the *Discrete Fourier Transform*:

$$F_k = \frac{1}{n} \sum_{j=0}^{n-1} x_j e^{-i2\pi jk}. \tag{A.2}$$

In our case we always have a real-valued series: $x_j \in \mathbb{R}$, and this induces a further constraint, as only $n/2$ (n even) or $n/2+1$ (n odd) are now independent, with the further constraint of $F_{n/2} \in \mathbb{R}$ if n is odd. Coefficients F_k can be computed efficiently using Fast Fourier Transform (FFT) algorithms,and give the component of our series with respect to the kth harmonic. Since we're not interested in the relative phase differences between harmonics, we usually consider the intensities $|F_k|^2$. In Fig. 4.11 we precisely plot $|F|^2$ as a function of period, which is the inverse of the harmonic k (i.e., the wavelength).

A.4. R_0 from differential equations

We start from Eq. 1.4. Since we wish to find the threshold, we linearize them around the critical point $x_i = \bar{x}_i, y_\alpha = 0$, which is the disease-free state (dfs). By definition of critical point, $f_i(\bar{x}, 0) = g_\alpha(\bar{x}, 0) = 0$. The linearized system around the dfs has the following Jacobian matrix $\hat{J}$:

$$\hat{J} = \left(\begin{array}{c|c} \frac{\partial f_i}{\partial x_j}\Big|_{dfs} & \frac{\partial f_i}{\partial y_\beta}\Big|_{dfs} \\ \hline \frac{\partial g_\alpha}{\partial x_j}\Big|_{dfs} & \frac{\partial g_\alpha}{\partial y_\beta}\Big|_{dfs} \end{array} \right) \tag{A.3}$$

We can use some features of disease models to put some constraint on $\hat{J}$. $\frac{\partial g_\alpha}{\partial x_j}\Big|_{dfs} = 0$ always, as all terms in g_α must contain at least one infectious compartment, which is zero when evaluated on the dfs. Hence, $\hat{J}$ is block triangular, so we can evaluate the stability condition separately for the two diagonal blocks. Since we're interested only in the dynamics of the infectious compartments, we concentrate on the sub-Jacobian in the lower right block: $J_{\alpha\beta} = \frac{\partial g_\alpha}{\partial y_\beta}\Big|_{dfs}$.

As in Sec. 1.5.1, we split $g_\alpha(x, y)$ into a recovery part $g_\alpha^{(1)}(y)$, and an infection part $g_\alpha^{(2)}(x, y)$, due to some reasonable constraints on the equations for the infectious compartments:

- They contain diagonal and linear terms with strictly negative coupling constants. This represents sick people recovering. Positive (or zero) coupling constants, or non-diagonal terms would refer to unrealistic conditions, and would place the system always above threshold.

- They can contain quadratic terms. In each quadratic term at least one compartment must be infectious. This represent transmissions.

- There are no term of higher order than quadratic. Again this would refer to unrealistic interaction scenarios among individuals.

Now we can decompose the Jacobian in its linear part (diagonal), and quadratic part: $J = J^{(1)} + J^{(2)}$, with

$$\begin{cases} J_{\alpha\beta}^{(1)} = \frac{\partial g_\alpha^{(1)}}{\partial y_\beta}\Big|_{dfs}, \\ J_{\alpha\beta}^{(2)} = \frac{\partial g_\alpha^{(2)}}{\partial y_\beta}\Big|_{dfs}. \end{cases} \tag{A.4}$$

The threshold condition is given by the value of the transmissibilities for which $\det J = 0$. Given that $\tilde{J}^{(1)}$ is invertible as coupling constant are strictly negative, we define

$$R_0 = -\rho[J^{(2)}(J^{(1)})^{-1}],\tag{A.5}$$

and see that $\det(J) = 0$ is equivalent to $R_0 = 1$.

Acknowledgements

My deepest gratitude goes to my supervisor, Vittoria Colizza, who took care of me from my MSc internship up until the end of my PhD. I will never be able to express my gratitude for all the time and energy she has invested on me, guiding me both personally and professionally all along these years. If *I am a scientist*, I owe that to her. I wish to thank Chiara Poletto, who has been my example, guide and friend during these PhD years. I then wish to thank Luca Ferreri, the perfect working mate, and most importantly a friend. I want to mention all the team at Epicx lab, with whom I've had a lot of fruitful discussion, and a lot of fun: Livio Bioglio, Pietro Coletti, Davide Colombi, Alexandre Darbon, Giancarlo De Luca, Valérie Labonté. I then want to mention all the friends scattered throughout the World that share the same joys and worries of a young researchers, and in particular the *Yologozza* team: Alberto Antonioni, Federico Battiston, Jacopo Iacovacci, Elisa Omodei. A debt of gratitude goes to Giovanni Petri, not only an outstanding scientist, but a friend who could give me the right advice at the right time. Then I wish to thank all the people and friends that accompanied me during these years, the ones that are still there, as well as the ones that followed different paths.

I wish to thank my Doctoral School, led by Dominique Costagliola, the Pierre Louis Institute of Epidemiology, my *équipe* leader Pierre-Yves Boëlle, and Guy Thomas, for giving me the right environment to develop my research and education.

Finally, I'll mention the most important people in my life, my parents. They've always supported and stood beside me, and I hope to make them proud with this achievement.

Bibliography

[1] Eugenio Valdano, Luca Ferreri, Chiara Poletto, and Vittoria Colizza. Analytical Computation of the Epidemic Threshold on Temporal Networks. *Phys. Rev. X*, 5(2):21005, April 2015.

[2] Eugenio Valdano, Chiara Poletto, Armando Giovannini, Diana Palma, Lara Savini, and Vittoria Colizza. Predicting Epidemic Risk from Past Temporal Contact Data. *PLoS Comput Biol*, 11(3):e1004152, 2015.

[3] Eugenio Valdano, Chiara Poletto, and Vittoria Colizza. Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution. *arXiv*, 2015.

[4] WHO. Report. *www.who.int/mediacentre/factsheets/fs310/en/index1*, 2012.

[5] Li Liu, Hope L Johnson, Simon Cousens, Jamie Perin, Susana Scott, Joy E Lawn, Igor Rudan, Harry Campbell, Richard Cibulskis, Mengying Li, Colin Mathers, and Robert E Black. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *The Lancet*, 379(9832):2151–2161, August 2015.

[6] CDC Report. Antibiotic Resistance Threats in the United States. *www.cdc.gov/drugresistance/threat-report-2013/*, 2013.

[7] Cory W. Morin, Andrew C. Comrie, and Kacey Ernst. Climate and dengue transmission: Evidence and implications. *Environmental Health Perspectives*, 121(11-12):1264–1272, 2013.

[8] Angela McLean. *SARS a case study in emerging infections*. Oxford University Press, Oxford New York, 2005.

[9] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases: the SARS case study. *BMC medicine*, 5(1):34, January 2007.

[10] Dirk Brockmann and Dirk Helbing. The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science*, 342(6164):1337–1342, December 2013.

[11] Christophe Fraser, Christl A Donnelly, Simon Cauchemez, William P Hanage, Maria D Van Kerkhove, T Dï¿½irdre Hollingsworth, Jamie Griffin, Rebecca F Baggaley, Helen E Jenkins, Emily J Lyons, Thibaut Jombart, Wes R Hinsley, Nicholas C Grassly, Francois Balloux, Azra C Ghani, Neil M Ferguson, Andrew Rambaut, Oliver G Pybus, Hugo Lopez-Gatell, Celia M Alpuche-Aranda, Ietza Bojorquez Chapela, Ethel Palacios Zavala, Dulce Ma. Espejo Guevara, Francesco Checchi, Erika Garcia, Stephane Hugonnet, Cathy Roth, and The W H O Rapid Pandemic Assessment Collaboration. Pandemic Potential of a Strain of Influenza A (H1N1): Early Findings. *Science*, 324(5934):1557–1561, 2009.

[12] Antoine Flahault, Elisabeta Vergu, and Pierre-Yves Boëlle. Potential for a global dynamic of Influenza A (H1N1). *BMC infectious diseases*, 9(1):129, January 2009.

[13] Yang Yang, Jonathan D Sugimoto, M Elizabeth Halloran, Nicole E Basta, Dennis L Chao, Laura Matrajt, Gail Potter, Eben Kenah, and Ira M Longini. The Transmissibility and Control of Pandemic Influenza A (H1N1) Virus. *Science*, 326(5953):729–733, 2009.

[14] Duygu Balcan, Hao Hu, Bruno Goncalves, Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, Vittoria Colizza, and Alessandro Vespignani. Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a Monte Carlo likelihood analysis based on human mobility. *BMC medicine*, 7(1):45, January 2009.

[15] Michele Tizzoni, Paolo Bajardi, Chiara Poletto, José J Ramasco, Duygu Balcan, Bruno Gonçalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC medicine*, 10(1):165, January 2012.

[16] Ali M Zaki, Sander van Boheemen, Theo M Bestebroer, Albert D M E Osterhaus, and Ron A M Fouchier. Isolation of a Novel Coronavirus from a Man with Pneu-

monia in Saudi Arabia. *New England Journal of Medicine*, 367(19):1814–1820, October 2012.

[17] Romulus Breban, Julien Riou, and Arnaud Fontanet. Interhuman transmissibility of Middle East respiratory syndrome coronavirus: estimation of pandemic risk. *The Lancet*, 382(9893):694–699, August 2013.

[18] Simon Cauchemez, Christophe Fraser, Maria D Van Kerkhove, Christl A Donnelly, Steven Riley, Andrew Rambaut, Vincent Enouf, Sylvie van der Werf, and Neil M Ferguson. Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. *The Lancet Infectious Diseases*, 14(1):50–56, August 2014.

[19] C. Poletto, C. Pelat, D. Lévy-Bruhl, Y. Yazdanpanah, P. Y. Boëlle, and V. Colizza. Assessment of the Middle East respiratory syndrome coronavirus (MERS-CoV) epidemic in the Middle East and risk of international spread using a novel maximum likelihood analysis approach. *Eurosurveillance*, 19(23), 2014.

[20] WHO Ebola Response Team. Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections. *New England Journal of Medicine*, 371(16):1481–1495, September 2014.

[21] MFC Gomes, A Pastore y Piontti, L Rossi, D Chao, I Longini, ME Halloran, and A Vespignani. Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak. *PLOS Currents Outbreaks*, 2014.

[22] Chiara Poletto, MF Gomes, Ana Pastore y Piontti, Luca Rossi, Livio Bioglio, Dennis L Chao, Ira M. Longini, M Elizabeth Halloran, Vittoria Colizza, and Alessandro Vespignani. Assessing the impact of travel restrictions on international spread of the 2014 West African Ebola epidemic. *Eurosurveillance*, 19(42):20936, 2014.

[23] Ousmane Faye, Pierre-Yves Boëlle, Emmanuel Heleze, Oumar Faye, Cheikh Loucoubar, N'Faly Magassouba, Barré Soropogui, Sakoba Keita, Tata Gakou, El Hadji Ibrahima Bah, Lamine Koivogui, Amadou Alpha Sall, and Simon Cauchemez. Chains of transmission and control of Ebola virus disease in Conakry, Guinea, in 2014: an observational study. *The Lancet Infectious Diseases*, 15(3):320–326, August 2015.

[24] W O Kermack and A G McKendrick. Contributions to the mathematical theory of epidemics - I. *Bulletin of Mathematical Biology*, 53(1-2):33–55, 1991.

[25] W O Kermack and A G McKendrick. Contributions to the mathematical theory of epidemics - II. The problem of endemicity. *Bulletin of Mathematical Biology*, 53(1-2):57–87, 1991.

[26] W O Kermack and A G McKendrick. Contributions to the mathematical theory of epidemics - III. Further studies of the problem of endemicity. *Bulletin of Mathematical Biology*, 53(1-2):89–118, 1991.

[27] Norman Bailey. *The mathematical theory of infectious diseases and its applications.* Griffin, London, 1975.

[28] R M Anderson and Robert M May. *Infectious Diseases of Humans: Dynamics and Control.* Oxford University Press, 1992.

[29] Daryl Daley. *Epidemic modelling : an introduction.* Cambridge University Press, Cambridge New York, 1999.

[30] Matt J. Keeling and P. Rohani. Modeling Infectious Diseases in Humans and Animals. *Princeton University Press*, 2007.

[31] Emilia Vynnycky. *An introduction to infectious disease modelling.* Oxford University Press, Oxford, 2010.

[32] Ciro Cattuto, Wouter den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks. *PLoS ONE*, 5(7):e11596, 2010.

[33] Marcel Salathé and James H Jones. Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Comput Biol*, 6(4):e1000736, 2010.

[34] Mohammad S Hashemian, Kevin G Stanley, and Nathaniel Osgood. FLUNET: Automated Tracking of Contacts During Flu Season. In *WiOpt'10: Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks*, pages 557–562, Avignon, France, May 2010.

[35] Marcel Salathé, Maria Kazandjieva, Jung Woo Lee, Philip Levis, Marcus W Feldman, and James H Jones. A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences*, 107(51):22020–22025, 2010.

[36] Lorenzo Isella, Mariateresa Romano, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Wouter Van den Broeck, Francesco Gesualdo, Elisabetta Pandolfi, Lucilla Ravà, Caterina Rizzo, and Alberto Eugenio Tozzi. Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors. *PLOS ONE*, 6(2):e17144, 2011.

[37] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology*, 271(1):166–180, 2011.

[38] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, and Philippe Vanhems. High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. *PLOS ONE*, 6(8):e23176, 2011.

[39] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Regis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, and Philippe Vanhems. Simulation of an SEIR Infectious Disease Model on the Dynamic Contact Network of Conference Attendees. *BMC Medicine*, 9(87), July 2011.

[40] Thomas Hornbeck, David Naylor, Alberto M Segre, Geb Thomas, Ted Herman, and Philip M Polgreen. Using Sensor Networks to Study the Effect of Peripatetic Healthcare Workers on the Spread of Hospital-Associated Infections. *The Journal of Infectious Diseases*, 206(10):1549–1557, November 2012.

[41] Wouter Van den Broeck, Marco Quaggiotto, Lorenzo Isella, Alain Barrat, and Ciro Cattuto. The making of Sixty-Nine Days Of Close Encounters At The Science Gallery. *Leonardo*, 45(3):285, May 2012.

[42] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, and Nicolas Voirin. Esti-

mating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. *PLoS ONE*, 8(9):e73970, 2013.

[43] Julie Fournet and Alain Barrat. Contact Patterns among High School Students. *PLoS ONE*, 9(9):e107878, 2014.

[44] Nicolas Voirin, Cécile Payet, Alain Barrat, Ciro Cattuto, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, Jean-Sébastien Casalegno, Bruno Lina, and Philippe Vanhems. Combining High-Resolution Contact Data with Virological Data to Investigate Influenza Transmission in a Tertiary Care Hospital. *Infection Control & Hospital Epidemiology*, 36(03):254–260, 2015.

[45] Thomas Obadia, Romain Silhol, Lulla Opatowski, Laura Temime, Judith Legrand, Anne C M Thiébaut, Jean-Louis Herrmann, Éric Fleury, Didier Guillemot, and Pierre-Yves Boëlle. Detailed Contact Data and the Dissemination of Staphylococcus aureus in Hospitals. *PLoS Comput Biol*, 11(3):e1004170, 2015.

[46] Marta C González, Cesar A Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, June 2008.

[47] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of Predictability in Human Mobility. *Science*, 327(5968):1018–1021, 2010.

[48] Gautier Krings, Márton Karsai, Sebastian Bernhardsson, Vincent D Blondel, and Jari Saramäki. Effects of time window size and placement on the structure of an aggregated communication network. *EPJ Data Science*, 1(1):1–16, 2012.

[49] Mikko Kivelä, Raj Kumar Pan, Kimmo Kaski, János Kertész, Jari Saramäki, and Márton Karsai. Multiscale analysis of spreading in a large communication network. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(03):P03005, 2012.

[50] Lauri Kovanen, Kimmo Kaski, János Kertész, and Jari Saramäki. Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proceedings of the National Academy of Sciences*, 110(45):18070–18075, 2013.

[51] Zhi-Qiang Jiang, Wen-Jie Xie, Ming-Xia Li, Boris Podobnik, Wei-Xing Zhou, and H Eugene Stanley. Calling patterns in human communication dynamics. *Proceedings of the National Academy of Sciences*, 110(5):1600–1605, 2013.

[52] Ming-Xia Li, Vasyl Palchykov, Zhi-Qiang Jiang, Kimmo Kaski, János Kertész, Salvatore Miccichè, Michele Tumminello, Wei-Xing Zhou, and Rosario N Mantegna. Statistically validated mobile communication networks: the evolution of motifs in European and Chinese data. *New Journal of Physics*, 16(8):83038, 2014.

[53] Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent D Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza. On the Use of Human Mobility Proxies for Modeling Epidemics. *PLoS Comput Biol*, 10(7):e1003716, 2014.

[54] Pablo Kaluza, Andrea Kölzsch, Michael T Gastner, and Bernd Blasius. The complex network of global cargo ship movements. *Journal of The Royal Society Interface*, January 2010.

[55] Lijun Sun, Kay W Axhausen, Der-Horng Lee, and Xianfeng Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34):13774–13779, 2013.

[56] Pierre Borgnat, Céline Robardet, Patrice Abry, Patrick Flandrin, Jean-Baptiste Rouquier, and Nicolas Tremblay. A Dynamical Network View of Lyon's V\'elo'v Shared Bicycle System. In Animesh Mukherjee, Monojit Choudhury, Fernando Peruani, Niloy Ganguly, and Bivas Mitra, editors, *Dynamics On and Of Complex Networks, Volume 2 SE - 13*, Modeling and Simulation in Science, Engineering and Technology, pages 267–284. Springer New York, 2013.

[57] Martin Rosvall, Alcides V Esquivel, Andrea Lancichinetti, Jevin D West, and Renaud Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nat Commun*, 5, August 2014.

[58] Ingo Scholtes, Nicolas Wider, René Pfitzner, Antonios Garas, Claudio J Tessone, and Frank Schweitzer. Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nat Commun*, 5, September 2014.

[59] Ken T D Eames and Matt J. Keeling. Monogamous networks and the spread of sexually transmitted diseases. *Mathematical biosciences*, 189(2):115–30, June 2004.

[60] Luis E C Rocha, Fredrik Liljeros, and Petter Holme. Information dynamics shape the sexual networks of Internet-mediated prostitution. *Proceedings of the National Academy of Sciences of the United States of America*, 107(13):5706–5711, 2010.

[61] R R Kao, Leon Danon, D M Green, and Istvan Z Kiss. Demographic structure and pathogen dynamics on the network of livestock movements in Great Britain. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1597):1999–2007, August 2006.

[62] Matthew C. Vernon and Matt J. Keeling. Representing the UK's cattle herd as static and dynamic networks. *Proceedings of the Royal Society of London B: Biological Sciences*, 276(1656):469–476, February 2009.

[63] Tom Lindström, Scott A Sisson, Susanna Stenberg Lewerin, and Uno Wennergren. Estimating animal movement contacts between holdings of different production types. *Preventive veterinary medicine*, 95(1-2):23–31, June 2010.

[64] Paolo Bajardi, Alain Barrat, Fabrizio Natale, Lara Savini, and Vittoria Colizza. Dynamical Patterns of Cattle Trade Movements. *PLoS ONE*, 6(5):e19869, 2011.

[65] Mario Konschake, Hartmut H K Lentz, Franz J Conraths, Philipp Hövel, and Thomas Selhorst. On the Robustness of In- and Out-Components in a Temporal Network. *PLoS ONE*, 8(2):e55223, 2013.

[66] Bhagat Lal Dutta, Pauline Ezanno, and Elisabeta Vergu. Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. *Preventive veterinary medicine*, 117(1):79–94, November 2014.

[67] Stephen Eubank, Hasan Guclu, V S Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, May 2004.

[68] Neil M Ferguson, Derek A T Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S Burke. Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature*, 437(7056):209–214, September 2005.

[69] Ira M Longini, Azhar Nizam, Shufu Xu, Kumnuan Ungchusak, Wanna Hanshaoworakul, Derek A T Cummings, and M Elizabeth Halloran. Containing Pandemic Influenza at the Source. *Science*, 309(5737):1083–1087, 2005.

[70] M Elizabeth Halloran, Neil M Ferguson, Stephen Eubank, Ira M Longini, Derek A T Cummings, Bryan Lewis, Shufu Xu, Christophe Fraser, Anil Vullikanti, Timothy C Germann, Diane Wagener, Richard Beckman, Kai Kadau, Chris Barrett, Catherine A Macken, Donald S Burke, and Philip Cooley. Modeling targeted layered containment of an influenza pandemic in the United States. *Proceedings of the National Academy of Sciences*, 105(12):4639–4644, 2008.

[71] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, 106(51):21484–21489, 2009.

[72] Dennis L Chao, M Elizabeth Halloran, Valerie J Obenchain, and Ira M Longini. FluTE, a Publicly Available Stochastic Influenza Epidemic Simulation Model. *PLoS Comput Biol*, 6(1):e1000656, January 2010.

[73] Stefano Merler, Marco Ajelli, Andrea Pugliese, and Neil M Ferguson. Determinants of the Spatiotemporal Dynamics of the 2009 H1N1 Pandemic in Europe: Implications for Real-Time Modelling. *PLoS Comput Biol*, 7(9):e1002205, September 2011.

[74] Hiroshi Nishiura. Prediction of pandemic influenza. *European Journal of Epidemiology*, 26(7):583–584, July 2011.

[75] Elaine O Nsoesie, John S Brownstein, Naren Ramakrishnan, and Madhav V Marathe. A systematic review of studies on forecasting the dynamics of influenza outbreaks. *Influenza and Other Respiratory Viruses*, 8(3):309–316, 2014.

[76] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[77] S N Dorogovtsev, A V Goltsev, and J F F Mendes. Critical phenomena in complex networks. *Rev. Mod. Phys.*, 80(4):1275–1335, October 2008.

[78] Carter T Butts. Revisiting the Foundations of Network Analysis. *Science*, 325(5939):414–416, 2009.

[79] Alessandro Vespignani. Predicting the Behavior of Techno-Social Systems. *Science*, 325(5939):425–428, 2009.

[80] M E J Newman. *Networks : an introduction.* Oxford University Press, Oxford New York, 2010.

[81] Chapter 12 - An Overview of Social Networks and Economic Applications. In *Handbook of Social Economics*, volume 1, pages 511–585. 2011.

[82] Alessandro Vespignani. Modelling dynamical processes in complex socio-technical systems. *Nat Phys*, 8(1):32–39, January 2012.

[83] Albert-László Barabási. *Network science.* Cambdridge University Press, Cambridge UK, 2015.

[84] Gary Chartrand. *Introductory graph theory.* Dover, New York, 1985.

[85] Frank Harary. *Graph theory.* Perseus Books, Reading, Mass, 1999.

[86] J A Bondy. *Graph theory.* Springer, New York, 2008.

[87] Alexei Vazquez, Balázs Rácz, András Lukács, and Albert-László Barabási. Impact of Non-Poissonian Activity Patterns on Spreading Processes. *Phys. Rev. Lett.*, 98(15):158702, April 2007.

[88] José Luis Iribarren and Esteban Moro. Impact of Human Activity Patterns on the Dynamics of Information Diffusion. *Phys. Rev. Lett.*, 103(3):38702, July 2009.

[89] Márton Karsai, Mikko Kivelä, R K Pan, K Kaski, J Kertész, Albert-László Barabási, and Jari Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E*, 83(2):25102, February 2011.

[90] Giovanna Miritello, Esteban Moro, and Rubén Lara. Dynamical strength of social ties in information spreading. *Phys. Rev. E*, 83(4):45102, April 2011.

[91] Luis E C Rocha and Vincent D Blondel. Bursts of Vertex Activation and Epidemics in Evolving Networks. *PLoS Comput Biol*, 9(3):e1002974, 2013.

[92] Luca Ferreri, Paolo Bajardi, Mario Giacobini, Silvia Perazzo, and Ezio Venturino. Interplay of network dynamics and heterogeneity of ties on spreading dynamics. *Phys. Rev. E*, 90(1):12812, July 2014.

[93] Luis E C Rocha, Fredrik Liljeros, and Petter Holme. Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts. *PLoS Comput Biol*, 7(3):e1001109, 2011.

[94] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, October 2012.

[95] Petter Holme and Jari Saramäki, editors. *Temporal networks*. Springer, Berlin New York, 2013.

[96] Petter Holme. Modern temporal network theory: A colloquium. *arXiv*, 2015.

[97] Thilo Gross, Carlos J Dommar D'Lima, and Bernd Blasius. Epidemic Dynamics on an Adaptive Network. *Phys. Rev. Lett.*, 96(20):208701, May 2006.

[98] Erik Volz and Lauren Ancel Meyers. Epidemic thresholds in dynamic contact networks. *Journal of The Royal Society Interface*, 6(32):233–241, March 2009.

[99] Michael Taylor, Timothy J Taylor, and Istvan Z Kiss. Epidemic threshold and control in a dynamic network. *Phys. Rev. E*, 85(1):16103, January 2012.

[100] Nicola Perra, Bruno Gonçalves, Romualdo Pastor-Satorras, and Alessandro Vespignani. Activity driven modeling of time varying networks. *Sci. Rep.*, 2, June 2012.

[101] Suyu Liu, Nicola Perra, Márton Karsai, and Alessandro Vespignani. Controlling Contagion Processes in Activity Driven Networks. *Phys. Rev. Lett.*, 112(11):118702, March 2014.

[102] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A Porter, Sergio Gómez, and Alexandre Arenas. Mathematical Formulation of Multilayer Networks. *Phys. Rev. X*, 3(4):41022, December 2013.

[103] Mikko Kivelä, Alexandre Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, September 2014.

[104] Nicholas C Grassly and Christophe Fraser. Mathematical models of infectious disease transmission. *Nat Rev Micro*, 6(6):477–487, June 2008.

[105] O Diekmann, J A P Heesterbeek, and J A J Metz. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology*, 28(4):365–382, 1990.

[106] Matt J. Keeling, Mark E J Woolhouse, Darren J Shaw, Louise Matthews, Margo Chase-Topping, Dan T Haydon, Stephen J Cornell, Jens Kappey, John Wilesmith, and Bryan T Grenfell. Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. *Science*, 294(5543):813–817, 2001.

[107] Leonid A. Rvachev and Ira M. Longini. A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75(1):3–22, July 1985.

[108] Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba. Epidemics with two levels of mixing. *Ann. Appl. Probab.*, 7(1):46–89, 1997.

[109] RebeccaF. Grais, J Hugh Ellis, and GregoryE. Glass. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European Journal of Epidemiology*, 18(11):1065–1072, 2003.

[110] Antoine Flahault, Elisabeta Vergu, Laurent Coudeville, and Rebecca F Grais. Strategies for containing a global influenza pandemic. *Vaccine*, 24(44-46):6751–5, December 2006.

[111] Vittoria Colizza, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani. Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions. *PLoS Med*, 4(1):e13, 2007.

[112] Robert Bartoszynski. Branching processes and the theory of epidemics. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Biology and Problems of Health*, pages 259–269, Berkeley, Calif., 1967. University of California Press.

[113] Peter Guttorp. *Stochastic modeling of scientific data*. Chapman & Hall, London New York, 1995.

[114] Christine Jacob. Branching processes: Their role in epidemiology. *International Journal of Environmental Research and Public Health*, 7(3):1186–1204, 2010.

[115] W John Edmunds, C J O'Callaghan, and D J Nokes. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proceedings of the Royal Society B: Biological Sciences*, 264(1384):949–957, July 1997.

[116] Philippe Beutels, Ziv Shkedy, Marc Aerts, and Pierre Van Damme. Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. *Epidemiology and Infection*, 134(6):1158–1166, December 2006.

[117] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael T Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, Janneke Heijne, Malgorzata Sadkowska-Todys, Magdalena Rosinska, and W John Edmunds. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med*, 5(3):e74, 2008.

[118] Rafael T Mikolajczyk, M K Akmatov, S Rastin, and M Kretzschmar. Social contacts of school children and the transmission of respiratory-spread pathogens. *Epidemiology & Infection*, 136(06):813–822, 2008.

[119] Niel Hens, Nele Goeyvaerts, Marc Aerts, Ziv Shkedy, Pierre Van Damme, and Philippe Beutels. Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC infectious diseases*, 9(1):5, January 2009.

[120] James M McCaw, Kristian Forbes, Paula M Nathan, Philippa E Pattison, Garry L. Robins, Terence M Nolan, and Jodie McVernon. Comparison of three methods for ascertainment of contact information relevant to respiratory pathogen transmission in encounter networks. *BMC infectious diseases*, 10(1):166, January 2010.

[121] Peter S Bearman, James Moody, and Katherine Stovel. Chains of affection: The structure of adolescent romantic and sexual networks. *American Journal of Sociology*, 110:44–91, 2002.

[122] Alan Neaigus, Samuel R Friedman, Marjorie Goldstein, Gilbert Ildefonso, Richard Curtis, and Benny Jose. Using dyadic data for a network analysis of HIV infection and risk behaviors among injecting drug users. *NIDA Res Monogr*, 151:20–37, 1995.

[123] David A. Rolls, Galina Daraganova, Rachel Sacks-Davis, Margaret Hellard, Rebecca Jenkinson, Emma McBryde, Philippa E Pattison, and G L Robins. Modelling hepatitis C transmission over a social network of injecting drug users. *Journal of theoretical biology*, 297:73–87, March 2012.

[124] David A. Rolls, Peng Wang, Rebecca Jenkinson, Philippa E Pattison, Garry L. Robins, Rachel Sacks-Davis, Galina Daraganova, Margaret Hellard, and Emma McBryde. Modelling a disease-relevant contact network of people who inject drugs. *Social Networks*, 35(4):699–710, October 2013.

[125] Leon Danon, Ashley P. Ford, Thomas House, Chris P. Jewell, Matt J. Keeling, Gareth O. Roberts, Joshua V. Ross, and Matthew C. Vernon. Networks and the epidemiology of infectious disease. *Interdisciplinary Perspectives on Infectious Diseases*, 2011, 2011.

[126] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *arXiv*, 2014.

[127] Norman Biggs. *Algebraic graph theory*. Cambridge University Press, Cambridge, 1993.

[128] Reuven Cohen. *Complex networks structure, robustness, and function*. Cambridge University Press, New York, 2010.

[129] Uwe Naumann. *Combinatorial scientific computing*. CRC Press, Boca Raton, 2012.

[130] A E Brouwer. *Spectra of graphs*. Springer, New York, NY, 2012.

[131] Albert-László Barabási and Réka Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509–512, 1999.

[132] Aaron Clauset, Cosma Rohilla Shalizi, and M E J Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703, 2009.

[133] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86(14):3200–3203, 2001.

[134] J O Lloyd-Smith, S J Schreiber, P E Kopp, and W M Getz. Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066):355–359, November 2005.

[135] Alison P Galvani and Robert M May. Epidemiology: Dimensions of superspreading. *Nature*, 438(7066):293–295, November 2005.

[136] Claudio Castellano and Romualdo Pastor-Satorras. Thresholds for Epidemic Spreading in Networks. *Phys. Rev. Lett.*, 105:218701, 2010.

[137] Paul E M Fine. Herd Immunity: History, Theory, Practice. *Epidemiologic Reviews*, 15(2):265–302, 1993.

[138] Paul E M Fine, Ken T D Eames, and David L Heymann. "Herd immunity": a rough guide. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 52(7):911–6, April 2011.

[139] Romualdo Pastor-Satorras and Alessandro Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65(3):36104, February 2002.

[140] Reuven Cohen, Shlomo Havlin, and Daniel Ben-Avraham. Efficient Immunization Strategies for Computer Networks and Populations. *Phys. Rev. Lett.*, 91(24):247901, December 2003.

[141] M E J Newman. Assortative Mixing in Networks. *Phys. Rev. Lett.*, 89(20):208701, October 2002.

[142] Marián Boguñá, Romualdo Pastor-Satorras, and Alessandro Vespignani. Absence of Epidemic Threshold in Scale-Free Networks with Degree Correlations. *Phys. Rev. Lett.*, 90(2):28701, January 2003.

[143] Tom Britton, Maria Deijfen, and Fredrik Liljeros. A Weighted Configuration Model and Inhomogeneous Epidemics. *Journal of Statistical Physics*, 145(5):1368–1384, 2011.

[144] Christel Kamp, Mathieu Moslonka-Lefebvre, and Samuel Alizon. Epidemic Spread on Weighted Networks. *PLoS Comput Biol*, 9(12):e1003352, 2013.

[145] P. van den Driessche and James Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical Biosciences*, 180(1-2):29–48, November 2002.

[146] M E J Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66(1):16128, 2002.

[147] David Chandler. *Introduction to modern statistical mechanics*. Oxford University Press, New York, 1987.

[148] Silvio C Ferreira, Claudio Castellano, and Romualdo Pastor-Satorras. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Phys. Rev. E*, 86(4):41125, October 2012.

[149] P Shu, W Wang, M Tang, and Y Do. Simulated identification of epidemic threshold on finite-size networks. *ArXiv e-prints*, October 2014.

[150] V I Arnold. *Ordinary differential equations*. MIT Press, Cambridge, 1973.

[151] Paul Glendinning. *Stability, instability, and chaos : an introduction to the theory of nonlinear differential equations*. Cambridge University Press, Cambridge England New York, 1994.

[152] O Diekmann. *Mathematical epidemiology of infectious diseases : model building, analysis, and interpretation*. John Wiley, Chichester New York, 2000.

[153] J M Heffernan, R J Smith, and L M Wahl. Perspectives on the basic reproductive ratio. *Journal of The Royal Society Interface*, 2(4):281–293, September 2005.

[154] Joaquï¿$\frac{1}{2}$n Marro. *Nonequilibrium phase transitions in lattice models*. Cambridge University Press, Cambridge New York, 2005.

[155] Yang Wang, D Chakrabarti, Chenxi Wang, and C Faloutsos. Epidemic spreading in real networks: an eigenvalue viewpoint. In *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, pages 25–34, 2003.

[156] Sergio Gómez, Alexandre Arenas, J Borge-Holthoefer, Sandro Meloni, and Yamir Moreno. Discrete time Markov chain approach to contact-based disease spreading in complex networks. *Europhysics Letters*, 89(3):38009, 2010.

[157] Marián Boguñá, Romualdo Pastor-Satorras, and Alessandro Vespignani. Epidemic Spreading in Complex Networks with Degree Correlations. In Romualdo Pastor-Satorras, Miguel Rubi, and Albert Diaz-Guilera, editors, *Statistical Mechanics of Complex Networks SE - 8*, volume 625 of *Lecture Notes in Physics*, pages 127–147. Springer Berlin Heidelberg, 2003.

[158] Reuven Cohen, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. Resilience of the Internet to Random Breakdowns. *Phys. Rev. Lett.*, 85(21):4626–4628, November 2000.

[159] Richard Durrett. *Probability : theory and examples*. Cambridge University Press, Cambridge New York, 2010.

[160] Ghassan Hamra, Richard MacLehose, and David Richardson. Markov chain Monte Carlo: an introduction for epidemiologists. *International journal of epidemiology*, 42(2):627–34, May 2013.

[161] Piet Van Mieghem. The N-intertwined SIS Epidemic Network Model. *Computing*, 93(2-4):147–169, 2011.

[162] Bernard C K Choi and Anita W P Pak. A Catalog of Biases in Questionnaires. *Prev Chronic Dis*, 2(1), 2005.

[163] Timo Smieszek, E U Burri, R Scherzinger, and R W Scholz. Collecting close-contact social mixing data with contact diaries: reporting errors and biases. *Epidemiology & Infection*, 140(04):744–752, 2012.

[164] Timo Smieszek, Victoria C Barclay, Indulaxmi Seeni, Jeanette J Rainey, Hongjiang Gao, Amra Uzicanin, and Marcel Salathé. How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC infectious diseases*, 14(1):136, January 2014.

[165] Fredrik Liljeros, Johan Giesecke, and Petter Holme. The Contact Network of Inpatients in a Regional Healthcare System. A Longitudinal Case Study. *Mathematical Population Studies*, 14(4):269–284, November 2007.

[166] Tjibbe Donker, Jacco Wallinga, and Hajo Grundmann. Patient Referral Patterns and the Spread of Hospital-Acquired Infections through National Health Care Networks. *PLoS Comput Biol*, 6(3):e1000715, March 2010.

[167] A Sarah Walker, David W Eyre, David H Wyllie, Kate E Dingle, Rosalind M Harding, Lily O'Connor, David Griffiths, Ali Vaughan, John Finney, Mark H Wilcox, Derrick W Crook, and Tim E A Peto. Characterisation of ¡italic¿Clostridium difficile¡/italic¿ Hospital Ward-Based Transmission Using Extensive Epidemiological Data and Molecular Typing. *PLoS Med*, 9(2):e1001172, February 2012.

[168] M Carolyn Gates and Mark E J Woolhouse. Controlling infectious disease through the targeted manipulation of contact network structure. *Epidemics*, pages –, 2015.

[169] M Lahiri and T Y Berger-Wolf. Structure Prediction in Temporal Networks using Frequent Subgraphs. In *Computational Intelligence and Data Mining, 2007. CIDM 2007. IEEE Symposium on*, pages 35–42, March 2007.

[170] Rajmonda Sulo, Tanya Berger-Wolf, and Robert Grossman. Meaningful Selection of Temporal Resolution for Dynamic Networks. In *Proceedings of the Eighth Workshop on Mining and Learning with Graphs*, MLG '10, pages 127–136, New York, NY, USA, 2010. ACM.

[171] Shi Chen, Amiyaal Ilany, Brad J White, Michael W Sanderson, and Cristina Lanzas. Spatial-Temporal Dynamics of High-Resolution Animal Networks: What Can We Learn from Domestic Animals? *PLoS ONE*, 10(6):e0129253, June 2015.

[172] David Kempe, Jon Kleinberg, and Amit Kumar. Connectivity and Inference Problems for Temporal Networks. *Journal of Computer and System Sciences*, 64(4):820–842, June 2002.

[173] Márton Karsai, Nicola Perra, and Alessandro Vespignani. Time varying networks and the weakness of strong ties. *Sci. Rep.*, 4, February 2014.

[174] Petter Holme. Epidemiologically Optimal Static Networks from Temporal Network Data. *PLoS Comput Biol*, 9(7):e1003142, 2013.

[175] Eddie Cheng, Jerrold W Grossman, and Marc J Lipman. Time-stamped graphs and their associated influence digraphs. *Discrete Applied Mathematics*, 128(2-3):317–335, June 2003.

[176] John Whitbeck, Marcelo de Amorim, Vania Conan, and Jean-Loup Guillaume. Temporal Reachability Graphs. In *Proceedings of the 18th Annual International*

Conference on Mobile Computing and Networking, Mobicom '12, pages 377–388, New York, NY, USA, 2012. ACM.

[177] Vincenzo Nicosia, John Tang, Cecilia Mascolo, Mirco Musolesi, Giovanni Russo, and Vito Latora. Graph Metrics for Temporal Networks. In Petter Holme and Jari Saramäki, editors, *Temporal Networks SE - 2*, Understanding Complex Systems, pages 15–40. Springer Berlin Heidelberg, 2013.

[178] Vladimir Batagelj and Selena Praprotnik. An algebraic approach to temporal network analysis based on temporal quantities. *arXiv*, 2015.

[179] Hartmut H K Lentz, Thomas Selhorst, and Igor M Sokolov. Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks. *Phys. Rev. Lett.*, 110(11):118701, March 2013.

[180] Kevin Lewis, Jason Kaufman, Marta C González, Andreas Wimmer, and Nicholas Christakis. Tastes, ties, and time: A new social network dataset using Facebook.com. *Social Networks*, 30(4):330–342, 2008.

[181] Michael Szell, Renaud Lambiotte, and Stefan Thurner. Multirelational organization of large-scale social networks in an online world. *Proceedings of the National Academy of Sciences*, 107(31):13636–13641, 2010.

[182] Jimi Adams, James Moody, and Martina Morris. Sex, Drugs, and Race: How Behaviors Differentially Contribute to the Sexually Transmitted Infection Risk Network Structure. *American Journal of Public Health*, 103(2):322–329, December 2012.

[183] Peter J Mucha, Thomas Richardson, Kevin Macon, Mason A Porter, and Jukka-Pekka Onnela. Community structure in time-dependent, multiscale, and multiplex networks. *Science*, 329:277, 2010.

[184] Danielle S Bassett, Mason A Porter, Nicholas F Wymbs, Scott T Grafton, Jean M Carlson, and Peter J Mucha. Robust detection of dynamic community structure in networks. *Chaos (Woodbury, N.Y.)*, 23(1):013142, March 2013.

[185] Marta Sarzynska, Elizabeth A Leicht, Gerardo Chowell, and Mason A Porter. Null Models for Community Detection in Spatially-Embedded, Temporal Networks. *arXiv*, 2014.

[186] Giovanna Miritello, Rubén Lara, Manuel Cebrian, and Esteban Moro. Limited communication capacity unveils strategies for human interaction. *Sci. Rep.*, 3, June 2013.

[187] M Ciccolini, J Arends, H Grundmann, and AW Friedrich. O085: A network-based approach using intra-hospital patient transfers to identify high-risk wards during nosocomial outbreaks, June 2013.

[188] Michele Starnini, Andrea Baronchelli, and Romualdo Pastor-Satorras. Modeling Human Dynamics of Face-to-Face Interaction Networks. *Phys. Rev. Lett.*, 110(16):168701, April 2013.

[189] Alain Barrat, Ciro Cattuto, V Colizza, F Gesualdo, Lorenzo Isella, E Pandolfi, Jean-François Pinton, Lucilla Ravà, Caterina Rizzo, Mariateresa Romano, Juliette Stehlé, Alberto Eugenio Tozzi, and Wouter Van den Broeck. Empirical temporal networks of face-to-face human interactions. *The European Physical Journal Special Topics*, 222(6):1295–1309, 2013.

[190] Christian L Vestergaard, Mathieu Génois, and Alain Barrat. How memory generates heterogeneous dynamics in temporal networks. *Phys. Rev. E*, 90(4):42805, October 2014.

[191] Hang-Hyun Jo, Juan I Perotti, Kimmo Kaski, and János Kertész. Analytically Solvable Model of Spreading Dynamics with Non-Poissonian Processes. *Phys. Rev. X*, 4(1):11041, March 2014.

[192] Antoine Moinet, Michele Starnini, and Romualdo Pastor-Satorras. Burstiness and Aging in Social Temporal Networks. *Phys. Rev. Lett.*, 114(10):108701, March 2015.

[193] Paul Erd\H os and Alfréd Rényi. On the Evolution of Random Graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5(17), 1960.

[194] R Milo, S Shen-Orr, S Itzkovitz, N Kashtan, D Chklovskii, and U Alon. Network Motifs: Simple Building Blocks of Complex Networks. *Science*, 298(5594):824–827, 2002.

[195] Emanuele Cozzo, Raquel A Baños, Sandro Meloni, and Yamir Moreno. Contact-based social contagion in multiplex networks. *Phys. Rev. E*, 88(5):50801, November 2013.

[196] Huijuan Wang, Qian Li, Gregorio D'Agostino, Shlomo Havlin, H Eugene Stanley, and Piet Van Mieghem. Effect of the interconnected network structure on the epidemic threshold. *Phys. Rev. E*, 88(2):22801, August 2013.

[197] E T Whittaker. *A course of modern analysis*. Cambridge University Press, Cambridge New York, 2000.

[198] Gerd Grubb. *Distributions and operators*. Springer, New York, 2009.

[199] Iain Anderson. Foot and Mouth Disease 2001: Lessons to be Learned Inquiry Report. *The Stationery Office*, (July), 2002.

[200] T. E. Carpenter, J. M. O'Brien, A. D. Hagerman, and B. A. McCarl. Epidemic and Economic Impacts of Delayed Detection of Foot-And-Mouth Disease: A Case Study of a Simulated Outbreak in California. *Journal of Veterinary Diagnostic Investigation*, 23(1):26–33, January 2011.

[201] L H Taylor, S M Latham, and Mark E J Woolhouse. Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1411):983–989, July 2001.

[202] Matt J. Keeling. Models of foot-and-mouth disease. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1569):1195–1202, June 2005.

[203] A-F Viet, C Fourichon, and H Seegers. Review and critical discussion of assumptions and modelling options to study the spread of the bovine viral diarrhoea virus (BVDV) within a cattle herd. *Epidemiology and infection*, 135(5):706–721, 2007.

[204] Aurélie Courcoul and Pauline Ezanno. Modelling the spread of Bovine Viral Diarrhoea Virus (BVDV) in a managed metapopulation of cattle herds. *Veterinary Microbiology*, 142(1-2):119–128, 2010.

[205] Mark Tinsley, Fraser I Lewis, and Franz Brülisauer. Network modeling of BVD transmission. *Veterinary research*, 43(1):11, January 2012.

[206] L.M. O'Reilly and C.J. Daborn. The epidemiology of Mycobacterium bovis infections in animals and man: A review. *Tubercle and Lung Disease*, 76:1–46, August 1995.

[207] Robin A. Skuce, Adrian R. Allen, and Stanley W. J. McDowell. Herd-level risk factors for bovine tuberculosis: A literature review. *Veterinary Medicine International*, 2012:621210, 2012.

[208] Ellen Brooks-Pollock, Gareth O. Roberts, and Matt J. Keeling. A dynamic model of bovine tuberculosis spread and control in Great Britain. *Nature*, 511(7508):228–231, July 2014.

[209] V Taleski, L Zerva, T Kantardjiev, Z Cvetnic, M Erski-Biljic, B Nikolovski, J Bosnjakovski, V Katalinic-Jankovic, A Panteliadou, S Stojkoski, and T Kirandziski. An overview of the epidemiology and epizootology of brucellosis in selected countries of Central and Southeast Europe. *Veterinary Microbiology*, 90(1-4):147–155, December 2002.

[210] T England, L Kelly, R D Jones, A MacMillan, and M Wooldridge. A simulation model of brucellosis spread in British cattle under several testing regimes. *Preventive veterinary medicine*, 63(1-2):63–73, April 2004.

[211] E Díaz Aparicio. Epidemiology of brucellosis in domestic animals caused by Brucella melitensis, Brucella suis and Brucella abortus. *Revue scientifique et technique (International Office of Epizootics)*, 32(1):43–51, 53–60, 2013.

[212] Claude Saegerman, Dirk Berkvens, and Philip S Mellor. Bluetongue Epidemiology in the European Union. *Emerging Infectious Disease journal*, 14(4):539, 2008.

[213] V Moennig, G Floegel-Niesmann, and I Greiser-Wilke. Clinical Signs and Epidemiology of Classical Swine Fever: A Review of New Knowledge. *The Veterinary Journal*, 165(1):11–20, January 2003.

[214] S Costard, L Mur, J Lubroth, J M Sanchez-Vizcaino, and D U Pfeiffer. Epidemiology of African swine fever virus. *Virus research*, 173(1):191–7, April 2013.

[215] Fabrizio Natale, Armando Giovannini, Lara Savini, Diana Palma, Luigi Possenti, Gianluca Fiore, and Paolo Calistri. Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread. *Preventive veterinary medicine*, 92(4):341–50, December 2009.

[216] Fabrizio Natale, Lara Savini, Armando Giovannini, Paolo Calistri, Luca Candeloro, and Gianluca Fiore. Evaluation of risk and vulnerability using a Disease

Flow Centrality measure in dynamic cattle trade networks. *Preventive veterinary medicine*, 98(2-3):111–8, February 2011.

[217] Paolo Bajardi, Alain Barrat, Lara Savini, and Vittoria Colizza. Optimizing surveillance for livestock disease spreading through animal movements. *Journal of The Royal Society Interface*, June 2012.

[218] Rebecca E LaFond and Sheila A Lukehart. Biological Basis for Syphilis. *Clinical Microbiology Reviews*, 19(1):29–49, January 2006.

[219] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms*, 6(2-3):161–180, 1995.