

Origin of Peer Influence in Social Networks

Flávio L. Pinheiro,^{1,2,3} Marta D. Santos,⁴ Francisco C. Santos,^{3,1} and Jorge M. Pacheco^{5,6,1}

¹*ATP Group, CMAF, Instituto para a Investigação Interdisciplinar, P-1649-003 Lisboa, Portugal*

²*Centro de Física da Universidade do Minho, 4710-057 Braga, Portugal*

³*INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, 2744-016 Porto Salvo, Portugal*

⁴*Departamento de Física and I3N, Universidade de Aveiro, 3810-193 Aveiro, Portugal*

⁵*Departamento de Matemática e Aplicações da Universidade do Minho, 4710-057 Braga, Portugal*

⁶*Centro de Biologia Molecular e Ambiental da Universidade do Minho, 4710-057 Braga, Portugal*

(Received 6 May 2012; revised manuscript received 3 October 2013; published 6 March 2014)

Social networks pervade our everyday lives: we interact, influence, and are influenced by our friends and acquaintances. With the advent of the World Wide Web, large amounts of data on social networks have become available, allowing the quantitative analysis of the distribution of information on them, including behavioral traits and fads. Recent studies of correlations among members of a social network, who exhibit the same trait, have shown that individuals influence not only their direct contacts but also friends' friends, up to a network distance extending beyond their closest peers. Here, we show how such patterns of correlations between peers emerge in networked populations. We use standard models (yet reflecting intrinsically different mechanisms) of information spreading to argue that empirically observed patterns of correlation among peers emerge naturally from a wide range of dynamics, being essentially independent of the type of information, on how it spreads, and even on the class of underlying network that interconnects individuals. Finally, we show that the sparser and clustered the network, the more far reaching the influence of each individual will be.

DOI: [10.1103/PhysRevLett.112.098702](https://doi.org/10.1103/PhysRevLett.112.098702)

PACS numbers: 89.65.Ef, 87.23.Kg, 89.75.Fb, 89.75.Hc

Human societies are embedded in complex social networks on which information flow—associated with traits such as emotions, behaviors, ideas, or fads—is ubiquitous [1–12]. What determines the patterns observed has become extremely valuable, with applications extending to all areas of human activity. Several studies have focused on the role played by social networks on the spread of information between individuals, by making use of email and blog databases and online social networks such as Twitter [11,12] and Facebook [6,7]. Such empirical studies have shown how social networks affect the propagation of health issues [13,14], ideas [15], criminal behavior [16,17], economic decisions [18,19], school achievement [20], and cooperation [21,22], among other human traits.

In what concerns the correlation patterns observed, Fowler and Christakis [22–25] recently proposed a *three degree of influence* rule based on the statistical analysis of the Framingham Heart Study database, from which a social network was inferred. Correlations among individuals were analyzed for traits as diverse as smoking habits, alcohol consumption, loneliness, obesity, cooperation, or happiness. These correlations reflected the relative increase in probability—when compared with a random arrangement—that two individuals share the same trait as a function of the network distance, defined as the smallest number of links connecting those individuals in the network [see Fig. 1(a)]. They found that similar and nontrivial correlations emerge from distinct traits and persist in the period studied, suggesting the validity of a three degrees of

influence on social networks. In other words, not only our “friends” but also our friends' friends together with their friends exhibit a positive correlation of traits. More recently, analysis of cooperation on social networks of hunter-gatherers revealed a degree of influence of 2 [26].

Here, we investigate the degree of peer influence that emerges from different dynamical processes representative of a plethora of phenomena occurring in networked populations—the spread of cooperative strategies, opinions, and diseases. Individuals are assigned to nodes of a complex network, whereas links between them represent interactions. We show that, for each network class considered, different processes often lead to the same degrees of influence, suggesting that peer influence is insensitive to the process at stake. On the other hand, we find that simple topological properties of the underlying networks, such as the average connectivity ($\langle k \rangle$) and the clustering coefficient, ultimately determine the number of degrees of influence observed, which systematically falls between 3 and 2, in agreement with the results stemming from empirical analyses of correlations in present [22–25] and past [26] social networks.

We start by studying the evolutionary dynamics of cooperation [5,27], modeled as peer-to-peer interactions by means of the famous Prisoner's Dilemma (PD) metaphor [27]. The word dilemma in the PD stems from the decision conflict that occurs when 2 individuals must simultaneously decide whether to Cooperate (*C*) or to Defect (*D*) towards the other. The game returns $R = 1$ for mutual

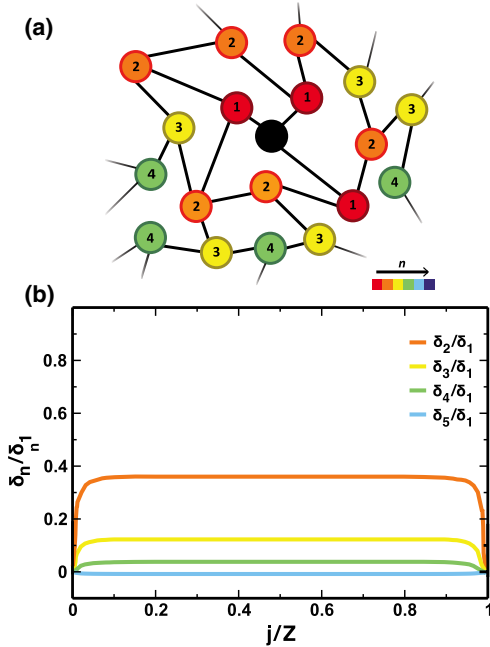


FIG. 1 (color online). (a) Network distance of a given node to the focal individual (black circle) defined as the smallest number of network links that separate the two (given by the number inside each circle). (b) Examples of correlation patterns as a function of j/Z (where j is the number of cooperators present in a population of size Z), emerging from the evolutionary PD game in homogeneous small world networks of size $Z = 10^3$ and degree $\langle k \rangle = 4$; the ratio $\delta_n(j/Z)/\delta_1(j/Z)$ provides an adequate normalization that renders the correlation patterns approximately constant for most values of j/Z . Similar patterns are obtained both for VM and SIR dynamics.

cooperation, $P = 0$ for mutual defection, $S = -\lambda$ when playing C against a D , and $T = 1 + \lambda$ when playing D against a C ($\lambda > 0$ measures both the temptation to defect and the fear of being cheated [28]). The ranking $T > R > P > S$ implies that maximizing one's own payoff leads each individual to choose D , irrespective of the decision of the other, such that the outcome will be mutual defection (pure Nash equilibrium for $\lambda > 0$). This outcome, however, is not the best for the pair, as mutual cooperation would lead to a better outcome for both of them. In the evolutionary version, the accumulated return from the interactions with all neighbors is interpreted as a measure of success (or “fitness”), such that some strategies (C or D) may become more attractive than others. We assume that individuals revise their behavior based on the perceived success of others: an individual A imitates a randomly chosen neighbor B with probability $p = [1 + e^{-\beta(f_B - f_A)}]^{-1}$, where f_A (f_B) stands for the fitness of A (B) and β denotes the intensity of selection [29]. In the mean-field limit, in which everyone is equally likely to interact with anyone else (also known as well-mixed population approximation), this dilemma inexorably condemns cooperation to extinction [27], a fate which may change when individuals are

embedded in a social network represented by means of a graph, in which structural diversity is ubiquitous [1–3,5,7–10,22–26,28,30–35].

In practice, however, the (population) dynamics of peer influence does not need to be fitness-driven or constrained by any type of social dilemma. To investigate opinion dynamics, we adopt a variant of one of the simplest and most commonly used models of opinion formation—the Voter Model (VM)—in which an individual can be influenced by a randomly chosen neighbor to adopt the opinion of the latter with probability $p = 1.0$ [10,36–38]. We confirmed that lower values of p do not change the results discussed below.

Moreover, information transmission has been sometimes regarded as contagious [15,39], similar to the propagation of infectious diseases [10]. Hence, we also study a widely used model of disease spreading, the Susceptible-Infected-Recovered model (SIR) [40], in which individuals can be in a susceptible (S), infected (I), or recovered (R) epidemiological state [10,40]. An infected individual can either infect a susceptible neighbor at an infection rate α or recover and become immune at a recovery rate γ .

Given the structural diversity of social networks [2,3,30,34] in which some individuals interact and/or are taken as role models more and more often than others, we investigate the role played by the distribution of the number of first neighbors of each individual (the degree distribution [2]) and clustering [2,35] (a measure of the number of individuals' neighbors who are also neighbors of each other) in the emerging patterns of correlations. To this end, we employ four network classes [30,41] with increasing variance of the degree distribution and similar low levels of clustering: homogeneous and heterogeneous small-world networks and exponential and scale-free networks [2,3,30]. Homogeneous small-world (HoSW) networks were obtained by repeatedly swapping the ends of pairs of randomly chosen links of a regular ring. Scale-free networks were obtained combining growth and preferential attachment, following the model proposed by Barabási and Albert (BA) [2]. Exponential networks (EXP) were obtained by adopting the same algorithm, with preferential attachment replaced by random attachment [2]. Heterogeneous small-world (HeSW) networks were built adopting the limit $p = 1$ of the Watts-Strogatz model [35], in which all links are rewired. Results reported in the main text refer to networks with $Z = 10^3$ (see Fig. 2) nodes, of the same order of magnitude of those investigated in Refs. [22–25] and for which finite-size effects are non-negligible. In the Supplemental Material [41], we investigate in detail the behavior of correlations as a function of the network size.

Let j be the number of individuals carrying one of the traits in a population of size Z . For each dynamical process and each j/Z we determine the propensity $\delta_n(j/Z)$ that 2 individuals at a distance n self-organize in the same trait

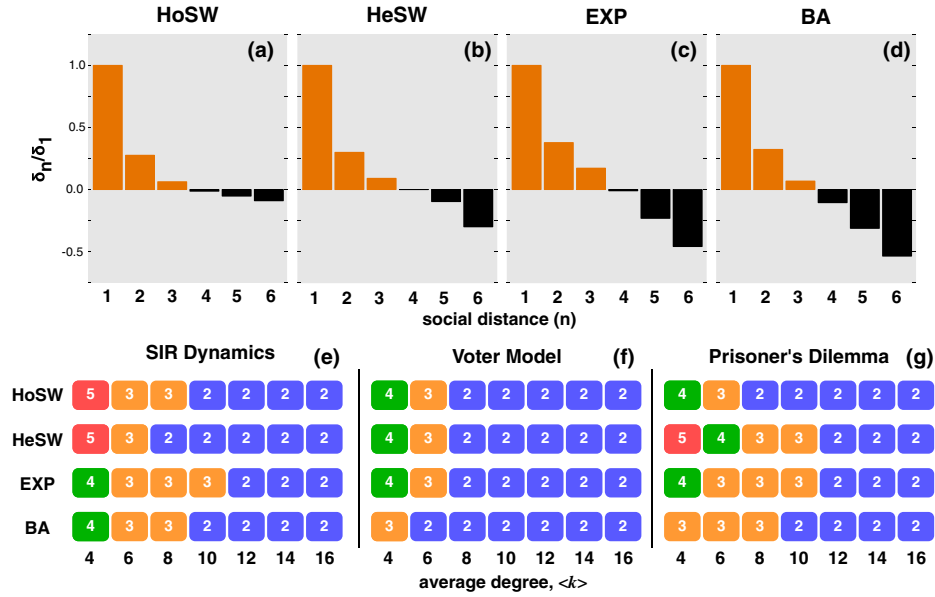


FIG. 2 (color online). Peer influence in social networks. Upper panels: δ_n/δ_1 for recovered and infected individuals in the SIR epidemic model taking place in (a) homogeneous small world networks, (b) heterogeneous small world networks, (c) exponential networks, and (d) scale-free networks. Lower panels: Critical degree of peer influence n_c (defined as the largest network distance n for which $\delta_n > 0$) for different values of average degree $\langle k \rangle$ and network classes shown in the top panels, for the following traits and processes (see main text for details). (e) Recovered and infected individuals in the SIR epidemic model (as above), (f) individuals with the same opinion in the VM, and (g) cooperators in the evolutionary PD game. All panels correspond to networks of size $Z = 10^3$, and results in the upper panels were obtained for networks with an average degree $\langle k \rangle = 6$. For larger network (and population) sizes, the negative values of the peer-influence correlation amplitudes that one obtains above n_c (see upper panels) will slowly approach 0, highlighting the importance of finite-size effects in evaluating peer-to-peer processes. The reason for the negative correlation values for $n > n_c$ in finite (and small) networks is clear: at a network distance n_c , most of the individuals with the same trait have been already sampled (see Fig. 2 in the Supplemental Material [41]); thus, for larger network distances, negative correlations will inevitably build up. At any rate, it is possible to clearly identify two regions of influence separated by n_c (see the Supplemental Material [41] for additional details).

relative to a random distribution of traits [23,24]. This quantity can be written as $\delta_n(j/Z) = \epsilon_n(j/Z)/\epsilon_n^{\text{rand}}(j/Z) - 1$, where $\epsilon_n(j/Z)$ is the average probability that a node shares the same trait with nodes located at a network distance n and $\epsilon_n^{\text{rand}}(j/Z)$ is the same quantity associated with a random distribution of traits, given by j/Z . The network distance (n) between individuals corresponds to the smallest number of links separating two nodes of a complex network, as depicted in Fig. 1(a). Additionally, we define n_c as the largest number of links (network distance) n for which δ_n remains positive ($n_c = 3$ in Refs. [22–25] and $n_c = 2$ in Ref. [26]).

To obtain $\epsilon_n(j/Z)$ we compute, for each network configuration, the average fraction of nodes that exhibit the same trait at a distance n . Hence, for each dynamical process, $\epsilon_n(j/Z)$ results from averaging over 10^6 independent network configurations. For the PD and VM processes, simulations were carried out starting from random configurations of traits; configurations included in the average were extracted after a transient of 10^3 generations (1 generation = Z iterations). For the SIR process, each simulation was started with a single I in a population of S ; α

and γ were defined such that the population would often reach a state with no I s left (Supplemental Material [41]).

The upper panels of Fig. 2 illustrate the normalized correlation values δ_n/δ_1 , which as shown in Fig. 1(b) are approximately independent from j/Z , obtained for the SIR dynamics in HoSW, HeSW, EXP, and BA networks. We observe that, regardless of the network topology, $n_c = 3$ as in Refs. [22–26]. Similar trends are also obtained for the two other dynamical processes introduced above (Supplemental Material [41]). This is shown in the lower panels of Fig. 2, which also show how the critical network distance (n_c) stabilizes at $n_c = 2$ with increasing network connectivity (e.g., $\langle k \rangle$), independently of the network structure, cluster coefficient, and dynamical process (see also the Supplemental Material [41]). On the contrary, deviations from this universal behavior of n_c are obtained only whenever networks become very sparse, associated with the smallest values of the average connectivity. Counterintuitively, the sparser the underlying network, the more far reaching the influence of each individual will be, extending beyond the most pervasive value of 2. Needless to say, the absolute values of the correlations, depicted in the

upper panels, will depend on the specific model parameters but not the final value of n_c .

Besides being heterogeneous, social networks often exhibit sizable levels of clustering [2], contrary to the negligible values that characterize the ones utilized in Fig. 2. To evaluate the impact of this property, we generate in the Supplemental Material [41] networks with arbitrary clustering for each network class [41,42]. We show how n_c remains limited between 2 and 4 for each of the three heterogeneous networks under study, irrespectively of their clustering coefficient and average degree. Nonetheless, we observe that increasing levels of clustering act to enlarge n_c , mostly whenever networks are very sparse, thus sizeably increasing their average path length (Supplemental Material [41]).

Overall, our results suggest that the extent of peer influence emerges as a natural outcome of dynamical processes on structured populations, being pervasive in a wide range of phenomena occurring in social networks. Despite the importance of social networks in defining the paths and ends of the dynamical processes they support, showing how important it is to address and understand population dynamics from a complex networks perspective [5,10,32,33,37], the patterns of peer influence they exhibit are surprisingly independent of their structure. On the other hand, when networks are very sparse, different network properties may contribute to enlarge the sphere of influence of each individual. Our results also show how networks naturally entangle individuals into interactions of many-body nature: Indeed, social networks effectively extend, in nontrivial ways, the dyadic interactions we started from. The fact that the network distance between any two individuals in social networks is small [2,3] and comparable to n_c further enhances the significance of the present results, as they stress how our individual actions may have wide scopes and counterintuitive repercussions.

Financial support by FEDER through POFC—COMPETE and by FCT-Portugal is gratefully acknowledged through Grants No. SFRH/BD/77389/2011, No. SFRH/BPD/90936/2012, No. PTDC/MAT/122897/2010, No. EXPL/EEI-SII/2556/2013, No. PESt-OE/EEI/LA0021/2013, and No. PESt-OE/BIA/UI4050/2014.

-
- [1] A. L. Lloyd and R. M. May, *Science* **292**, 1316 (2001).
 [2] R. Albert and A. L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
 [3] S. N. Dorogotsev and J. F. F. Mendes, *Evolution on Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, New York, 2003).
 [4] M. Rosvall and K. Sneppen, *Phys. Rev. Lett.* **91**, 178701 (2003).
 [5] G. Szabó and G. Fáth, *Phys. Rep.* **446**, 97 (2007).
 [6] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, *Soc. Networks* **30**, 330 (2008).

- [7] J. P. Onnela and F. Reed-Tsochas, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18 375 (2010).
 [8] D. Liben-Nowell and J. Kleinberg, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4633 (2008).
 [9] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, *Phys. Rep.* **469**, 93 (2008).
 [10] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).
 [11] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC* (2010).
 [12] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, in *Proceedings of the Fourth International Conference on Web Search and Data Mining, Hong-Kong* (2011), p. 65.
 [13] J. G. Trogdon, J. Nonnemaker, and J. Pais, *J. Health Econ.* **27**, 1388 (2008).
 [14] H. Nair, P. Manchanda, and T. Bhatia, *J. Market. Res.* **47**, 883 (2010).
 [15] L. Bettencourt, A. Cintron-Arias, D. Kaiser, and C. Castillo-Chavez, *Physica (Amsterdam)* **364A**, 513 (2006).
 [16] E. L. Glaeser, B. Sacerdote, and J. A. Scheinkman, *Q. J. Econ.* **111**, 507 (1996).
 [17] A. Calvó-Armengol and Y. Zenou, *Int. Econ. Rev.* **45**, 939 (2004).
 [18] A. Calvó-Armengol, *J. Econ. Theory* **115**, 191 (2004).
 [19] L. Cohen, A. Fazzini, and C. Malloy, *J. Polit. Econ.*, **116**, 951 (2008).
 [20] B. Sacerdote, *Q. J. Econ.* **116**, 681 (2001).
 [21] P. Brañas-Garza, R. Cobo-Reyes, M. Paz Espinosa, N. Jiménez, J. Kovářík, and G. Ponti, *Games Econ. Behav.* **69**, 249 (2010).
 [22] J. Fowler and N. Christakis, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5334 (2010).
 [23] N. Christakis, and J. Fowler, *N. Engl. J. Med.* **358**, 2249 (2008).
 [24] N. Christakis and J. Fowler, *N. Engl. J. Med.* **357**, 370 (2007).
 [25] J. Fowler and N. Christakis, *Br. Med. J.* **337**, a2338 (2008).
 [26] C. L. Apicella, F. W. Marlowe, J. H. Fowler, and N. A. Christakis, *Nature (London)* **481**, 497 (2012).
 [27] K. Sigmund, *The Calculus of Selfishness* (Princeton University Press, Princeton, NJ, 2010).
 [28] F. C. Santos, J. M. Pacheco, and T. Lenaerts, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3490 (2006).
 [29] G. Szabó and C. Tóke, *Phys. Rev. E* **58**, 69 (1998).
 [30] L. A. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11149 (2000).
 [31] F. C. Santos and J. M. Pacheco, *Phys. Rev. Lett.* **95**, 098104 (2005).
 [32] M. Perc and A. Szolnoki, *BioSystems* **99**, 109 (2010).
 [33] F. L. Pinheiro, J. M. Pacheco, and F. C. Santos, *PLoS One* **7**, e32114 (2012).
 [34] J. P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A.-L. Barabasi, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332 (2007).
 [35] D. Watts and S. Strogatz, *Nature (London)* **393**, 440 (1998).
 [36] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes* (Springer-Verlag, Berlin, 1999).
 [37] T. Antal, S. Redner, and V. Sood, *Phys. Rev. Lett.* **96**, 188104 (2006).

- [38] V. Sood, T. Antal, and S. Redner, *Phys. Rev. E* **77** 041121 (2008).
- [39] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, *Emotional Contagion* (Cambridge University Press, Cambridge, England, 1994).
- [40] W. O. Kermack and A. G. McKendrick, *Proc. R. Soc. A* **115**, 700 (1927).
- [41] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.112.098702> for additional details of the models and methods used, and for results obtained for larger network sizes.
- [42] S. Bansal, S. Khandelwal, and L. A. Meyers, *BMC Bioinf.* **10**, 405 (2009).

Origin of Peer Influence in Social Networks

Flávio L. Pinheiro^{1,2,3}, Marta D. Santos^{4,1}, Francisco C. Santos^{3,5,1} and Jorge M. Pacheco^{6,7,1}

¹ ATP-Group, CMAF, Instituto para a Investigação Interdisciplinar da Universidade de Lisboa, P-1649-003 Lisboa, Portugal,

² Centro de Física da Universidade do Minho, 4710 - 057 Braga, Portugal,

³ INESC-ID, IST-Taguspark, 2744-016 Porto Salvo, Portugal,

⁴ Departamento de Física & I3N, Universidade de Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

⁵ DEI, Instituto Superior Técnico, Universidade de Lisboa, IST- Taguspark, 2744-016 Porto Salvo, Portugal,

⁶ Departamento de Matemática e Aplicações da Universidade do Minho, 4710 - 057 Braga, Portugal

⁷ Centro de Biologia Molecular e Ambiental da Universidade do Minho, 4710 - 057 Braga, Portugal

Supplemental Material

Overview

This document provides *supplementary information* on the models and methodologies employed for the computation and analysis of the Peer Influence patterns discussed in the [Main text](#). The following list overviews the contents of this document:

- [Section 1](#) provides an overview of the topological features of the complex networks adopted in the main text, which are relevant for the analysis of peer-influence patterns;
- [Section 2](#) describes how peer influence patterns are defined along with how trait correlations were computed;
- [Section 3](#) provides results on the *Susceptible-Infected-Recovered epidemics dynamics*,
- [Section 4](#) provides results on the *Voter Model dynamics*,
- [Section 5](#) provides results on the *Prisoner's Dilemma dynamics*.

1. Social Structure

Let us consider a population of Z individuals embedded in a complex network of social interactions, where each node represents an individual and an undirected link a social interaction between two individuals. All links have the same weight equal to 1. The total number of interactions an individual i engages defines his degree k_i , while the fraction of individuals with degree k is given by the degree distribution $D(k)$. We shall refer to a network as *homogeneous* if all individuals have the same degree and *heterogeneous* otherwise. Moreover, we quantify the level of degree heterogeneity of a network by the variance of the degree distribution ($\sigma^2(k)$) [1].

Four different network topologies with different degree distributions are investigated in this work: *Homogeneous Small World (HoSW)* networks; *Heterogeneous Small World (HeSW)* networks; *Exponential (EXP)* and *scale-free (BA)* networks. Among these popular network types, **BA** and **EXP** networks are often regarded as good models of real world networks [2].

HoSW networks were generated by repeatedly swapping the ends of links from an initially regular network [3]. As a result all node-node correlations are eliminated while retaining their homogeneous character. **HeSW** networks were generated by rewiring half of the links of each node of a regular ring graph to a new random node in the population [4], forbidding link duplication and requiring overall network connectivity. The resulting degree distribution follows a *Poisson* distribution with k above $\langle k \rangle / 2$ [4], where $\langle k \rangle$ is the average connectivity of the network.

EXP networks are characterized by a degree distribution that follows an *geometric distribution* (exponential in the limit of very large populations) and were obtained by means of a growth mechanism with linear (random) attachment [5]. Similarly, **BA**

networks were constructed by means of a growth mechanism combined with preferential attachment, which generates *power law* degree distribution [5, 6] characterized by the existence of a small group of nodes with a large number interactions, while the majority of the population engages in few.

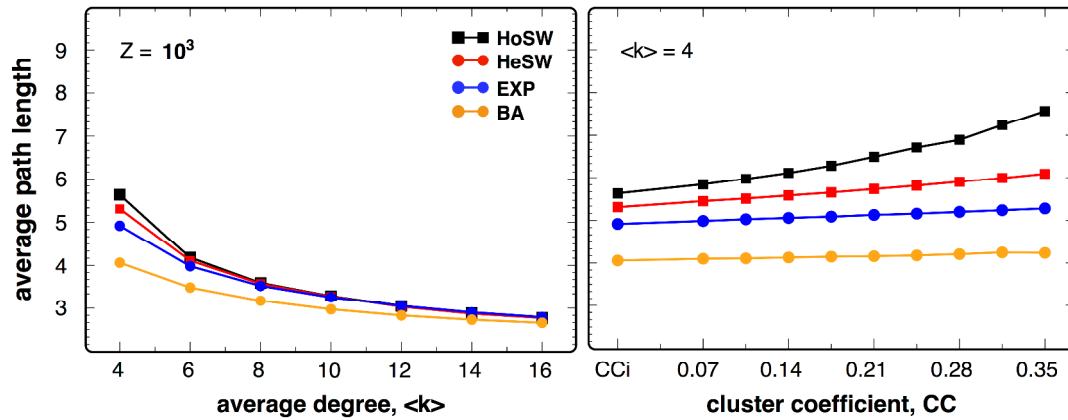


Figure SM.1 – *Average Path Length* as a function of the average degree (left panel) and cluster coefficient (right panel). Different colors are used for each of the topologies studied (as indicated). Each network has $Z = 10^3$ individuals; each data point corresponds to an average over 100 different networks.

Apart from the degree distribution, two other quantities of interest are used to characterize network topology: The *clustering coefficient* (CC) and the *average path length* (APL). The first concerns the fraction of neighboring individuals that share a common neighbor [7] and the second the average network distance (measured in terms of links traversed) between any two random individuals of the network [7].

By default, each network has a low CC (<0.035); thus in order to inspect the impact of increasing CC on each topology we employ an algorithm proposed in [8] that increases the CC without changing the degree distribution, and that works as follows: at each iteration the ends of two randomly selected links are swapped; if the resulting network has a higher CC we validate this rewiring; otherwise we return to the initial network. These steps are repeated until the desired CC is attained.

Figure SM.1 shows how the **APL** changes with the $\langle k \rangle$ (left panel) and **CC** (right panel) for each of the topologies under study ($Z = 10^3$). Naturally, for a fixed **CC**, increasing $\langle k \rangle$ leads to a decrease of the **APL**; on the other hand, increasing **CC** for fixed $\langle k \rangle$ leads to an increase in the **APL**.

Figure SM.2, in turn, shows the impact of network size on the average fraction of nodes at a distance of n links from any random node, a distribution that we denote as $\eta(n)$.

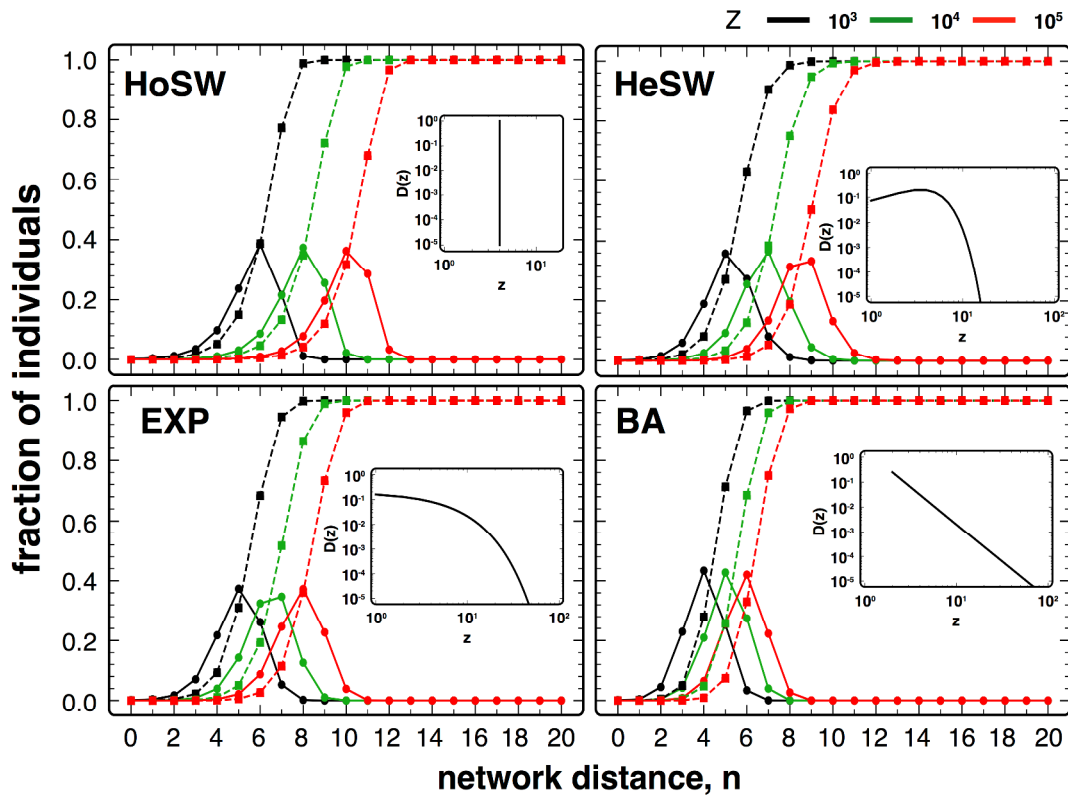


Figure SM.2 – Average fraction of nodes at distance n , $\eta(n)$ from a random node of the population. Each panel shows the results for a different type of network: *Homogeneous Small World networks* (upper left panel); *Heterogeneous Small World networks* (upper right panel); *exponential networks* (bottom left panel) and *scale-free networks* (bottom right panel). Dashed lines denote the accumulated distribution, that is the fraction of nodes at a network distance lower or equal to n . Different colors refer to networks with different size: $Z = 10^3$ (black); 10^4 (green) and 10^5 (red).

This network property provides a valuable indicator within this context, as it provides a measure of the sampling size available for each network distance and network topology (see Section 2). Interestingly, $\eta(n)$ peaks at different values of n for different population

sizes (Z), shifting to larger n with larger Z , allowing for better statistics to be obtained for larger values of n .

In sections 3, 4 and 5 we provide results of extensive numerical simulations in which we investigate the impact of different dynamical processes on the peer influence patterns of structured populations with $Z = 10^3, 10^4$ and 10^5 . This is done for different levels of *clustering coefficient* CC and average degree $\langle k \rangle$.

2. Peer Influence

Let us consider a population of Z individuals in which j share a common trait. We designate by trait configuration any of the $Z + 1$ available states of the population characterized by j , that is $0 \leq j/Z \leq 1$.

Let us define $\varepsilon_n(j/Z)$ as the probability that for a given j/Z two individuals at the distance of n links share the same trait. Therefore, if traits are distributed at random along the nodes of the population we expect $\varepsilon_n(j/Z)$ to be, on average, equal to j/Z . The existence of positive/negative correlations of traits between individuals at a distance of n links can be conveniently measured by computing $\varepsilon_n(j/Z) - j/Z$, where positive (negative) correlations of traits stand for individuals being assorted preferentially close to (away from) individuals with a similar trait.

For convenience, we also define $\delta_n(j/Z) \in [-1, 1]$ as the average correlation of traits (normalized) between two nodes at the distance of n links, that is

$$\delta_n(j/Z) = \frac{\varepsilon_n(j/Z) - j/Z}{j/Z} \quad (1)$$

where the division by j/Z provides the convenient normalization that allows an easy interpretation and comparison with other works in the literature [9].

The computation of $\delta_n(j/Z)$ requires the analysis of the network patterns of traits (that we refer to as *network configurations*) generated by each dynamical process under analysis. This task is divided in two steps: **(1)** a set of m *network* configurations produced by a dynamical process are extracted and **(2)** for each *network* configuration we compute $\delta_n(i;j/Z)$ (correlation at distance n for *network* configuration $1 \leq i \leq m$). The first step depends on the dynamical process at hand and hence the extraction of the *network* configurations is detailed at the beginning of sections 3, 4 and 5. As for the second step, $\delta_n(j/Z)$ is computed as the average of $\delta_n(i;j/Z)$ over all m samples while the error bars shown in Figures below correspond to the expected deviations from the average, that is, $\sigma_n(j/Z)/\sqrt{m}$.

The quantity of interest in our analysis is the critical network distance above which $\delta_n(j/Z)$ becomes statistically negative, n_c , that is, when $\delta_n(j/Z) - \sigma_n(j/Z)/\sqrt{m} < 0$.

3. SIR Dynamics

SIR models and extensions have been extremely successful in investigating the propagation of contagious diseases from which individuals are able to acquire immunity. Chickenpox, measles and seasonal flu are a few real life examples of diseases that fall in this category. Here we follow closely the model studied by *Segbroeck et al* [10].

Consider a population composed by individuals that can be, at any time, in one of three states: **S** (*Susceptible*); **I** (*Infected*) and **R** (*Recovered*), evolving under an asynchronous dynamics where at each *time-step* a *recovery* event occurs with probability $p = (1 + \langle z \rangle)^{-1}$ otherwise, with probability $1-p$, an *infection* event occurs. Additionally, at each time step an individual A is selected at random from the population; if A is in state **I** and a *recovery* event was selected then **I** becomes **R** with probability α ; however if A was in state **S** and an *infection* event was selected then A becomes **I** with probability

β if a randomly chosen neighbor of A is also in state I . For the purpose of computing correlations between traits, we consider correlations between S and both R and I , in the sense that the *network* pattern stems from the same dynamical event: S 's becoming I 's.

Each *network configuration* was extracted from a simulation that starts with a single source of infection. The outbreak evolved until the population reached the desired *trait configuration*. This procedure was repeated as many times as needed¹.

An appropriate β/α was chosen to ensure that most of the population becomes infected from a single infected individual regardless of the population structure, that is, above the most demanding epidemic threshold [11]. The peer influence patterns detected are, however, consistent for a wide range of β/α under this regime, thus providing a convenient set up in which to analyze the *peer influence* patterns that emerge for the entire domain of trait configurations.

Besides the results already shown in Fig. 2 of main text, Figure SM.3 shows how n_c changes with the CC (between 0 and 0.35, the range in which values obtained for real world networks fall) on heterogeneous populations.

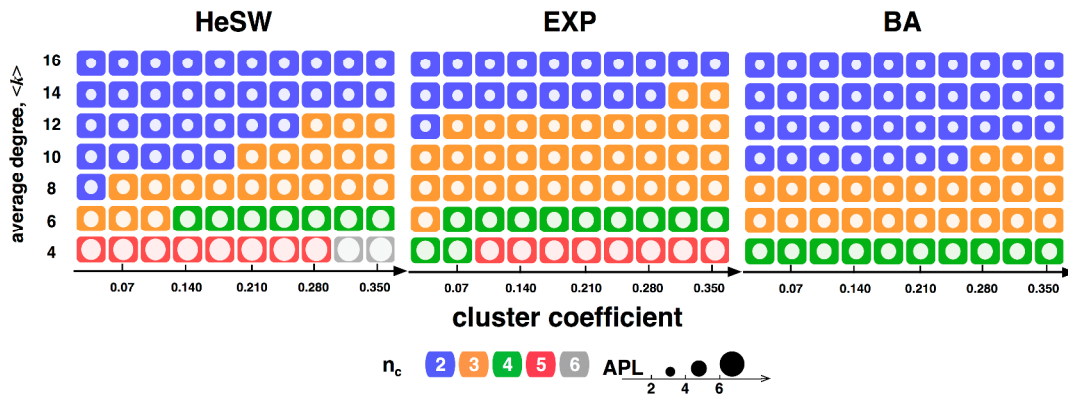


Figure SM.3 – Critical network distance on heterogeneous social networks with different average degrees and cluster coefficient. Populations have $Z = 10^3$ individuals, each colored cell denotes a different network and circles correspond to their relative *average path length* (APL), which decreases with increasing *average degree* ($\langle k \rangle$) and decreasing *cluster coefficient* (CC).

¹ Typically, 10^6 times for populations with $Z = 10^3$ individuals, 10^5 times for larger population sizes.

Figure SM.3 evidences how n_c increases with increasing **APL**. This feature is explained by the fact that the network diameter (the maximum distance between any two nodes) follows the same trend of the **APL**, thus increasing with increasing **CC** and decreasing $\langle k \rangle$.

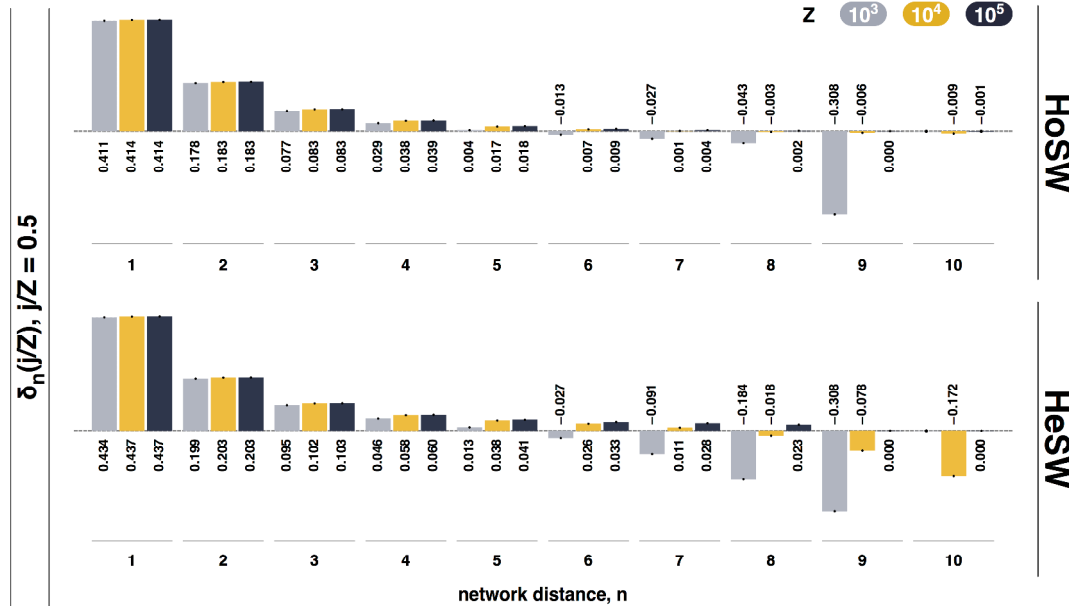


Figure SM.4 – Peer Influence patterns on *Homogeneous* (upper panel) and *Heterogeneous* (bottom panel) *Small World* populations with different sizes. We assume populations composed by $Z = 10^3$ (gray), 10^4 (yellow) and 10^5 (black) individuals and average degree $\langle k \rangle = 4$ and show how the correlation values vary with the network distance.

Figures SM.4 and SM.5 show how the impact of population size (10^3 , 10^4 and 10^5) on the peer influence patterns generated from the **SIR** dynamics, putting in evidence the role of finite size effects. Values in the figures below (above) the bars identify a positive (negative) $\delta_n(j/Z)$. The absence of values above/bellow a certain value of n implies that there are no nodes at a distance of n links for that particular network size. Finally, bars of different colors represent populations of different sizes (gray correspond to 10^3 , yellow to 10^4 and black to 10^5).

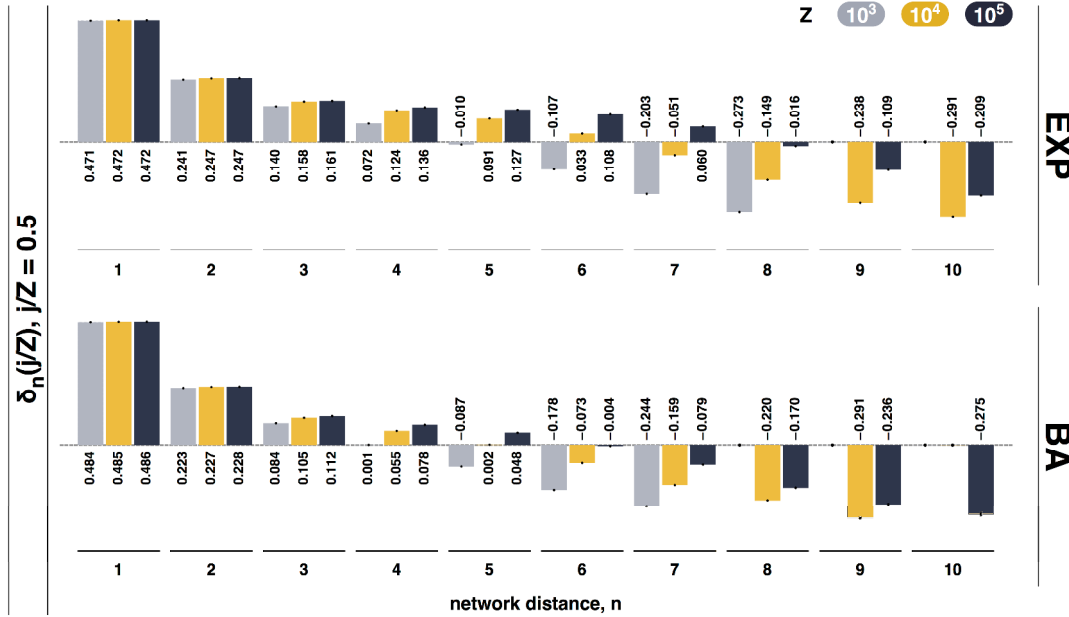


Figure SM.5 – Peer Influence patterns on *Exponential* (upper panel) and *scale free* (bottom panel) populations of different sizes. We assume populations composed by $Z = 10^3$ (gray), 10^4 (yellow) and 10^5 (black) individuals and average degree $\langle k \rangle = 4$ and show how the correlation values vary with the network distance.

For populations with 10^3 individuals, $\delta_n(j/Z)$ monotonically decreases with increasing n on all topologies. Thus there is a well defined *critical network distance* n_c above which $\delta_n(j/Z)$ become (increasingly) negative. For larger populations, however, this pattern is only kept on strongly heterogeneous populations (**EXP** and **BA**), while on *Small World* networks (**HoSW** and **HeSW**) $\delta_n(j/Z)$ tends to zero with increasing n , thus becoming difficult to identify n_c .

4. Voter Model

The Voter Model presents a simple yet highly successful framework in which to study opinion formation [12-15]. Here we assume a structured population of individuals. Each individual can have one of two possible traits $S_i = 0$ or 1. At each *time-step* (iteration) a randomly selected individual from the population imitates the trait of a randomly selected neighbor with probability p , which, without loss of generality, we consider equal to one.

Each *network configuration* was extracted from populations that evolve from an initially random configuration of traits for a period of $\tau_i = 1.5 \times 10^3$.

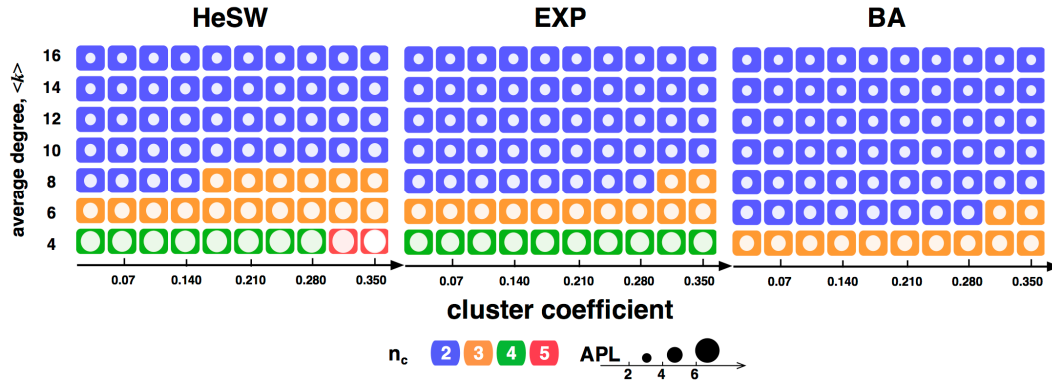


Figure SM.6 – Critical network distance on heterogeneous social networks with different average degree and cluster coefficient. Populations have $Z = 10^3$ individuals. Each colored cell denotes a different network and circles correspond to their *average path length (APL)*, which decreases with increasing *average degree* ($\langle k \rangle$) and decreasing *cluster coefficient (CC)*.

A transient of $\tau_i = 10^3$ generations was considered. After the transient regime, *network configurations* associated with the desired trait configuration (j^*/Z) were stored with probability $\sim 10^{-3}$, thus avoiding the same *network configuration* to be stored in consecutive *time-steps*.

Once again we found that the critical network distance, n_c , tends to increase with increasing **APL** (white disks), see [Figure SM.6](#). This happens for decreasing $\langle k \rangle$ or increasing **CC**. [Figures SM.7](#) and [SM.8](#) show the impact of population size (10^3 , 10^4 and 10^5) on the peer influence patterns generated from Voter Model dynamics.

Similar to the results obtained with the SIR dynamics, for populations with 10^3 , $\delta_i(j/Z)$ decreases monotonically with increasing network distance on all topologies. These results, however, contrast with the pattern of $\delta_i(j/Z)$ on larger populations that, overall (**BA** topologies constitute an exception) tend to zero with increasing network distance on all topologies.

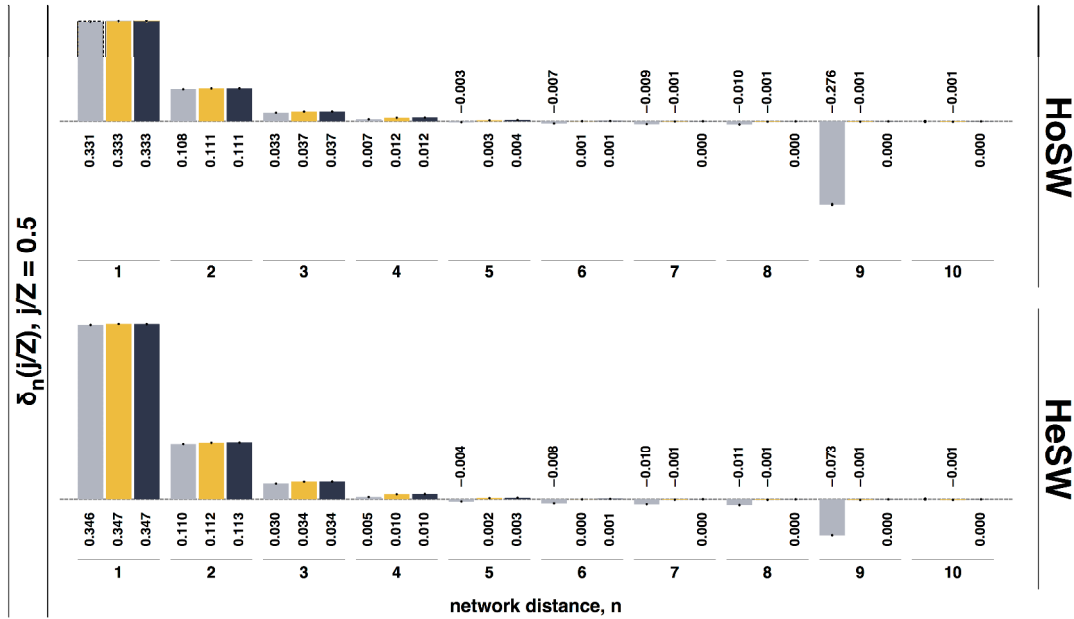


Figure SM.7 – Peer Influence patterns on *Homogeneous* (upper panel) and *Heterogeneous* (bottom panel) *Small World* populations with different sizes. We assume populations composed by $Z = 10^3$ (gray), 10^4 (yellow) and 10^5 (black) individuals and average degree $\langle k \rangle = 4$ and show how the correlation values vary with the network distance.

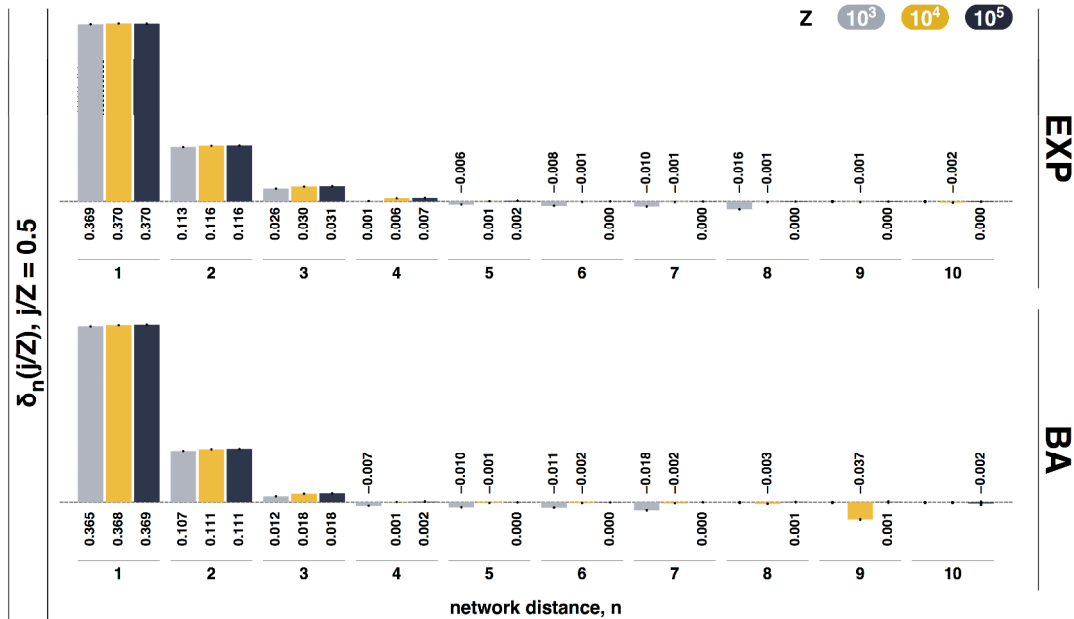


Figure SM.8 – Peer Influence patterns on *Exponential* (upper panel) and *scale free* (bottom panel) populations of different sizes. We assume populations composed by $Z = 10^3$ (gray), 10^4 (yellow) and 10^5 (black) individuals and average degree $\langle k \rangle = 4$ and show how the correlation values vary with the network distance.

5. Prisoner's Dilemma

Making use of *Evolutionary Game Theory* (EGT) [16-18] we now study the patterns of peer influence generated by a fitness driven evolutionary dynamics. Here we consider the most popular social dilemma of cooperation: The *Prisoner's Dilemma* (PD [18]).

Let us consider that each individual adopts unconditionally one of two strategies (to *Cooperate* or to *Defect*). Each interaction entails a payoff to both participating individuals that depends only on the pair strategy. If both individuals cooperate each earns the *reward* (**R**) for mutual cooperation; if both defect each gets the *punishment* (**P**) for mutual defection; finally a cooperator facing a defector obtains the *sucker's payoff* (**S**) for being cheated upon, whereas the defector obtains the so called *temptation* (**T**) payoff. Here we shall adopt a standard parameterization of these four payoffs, making **R** = 1.0, **P** = 0.0, **T** = 1 + λ and **S** = $-\lambda$; under these assumptions, $\lambda > 0$ ensures that individuals engage in a **PD**.

At each *time-step* a randomly selected individual **A** updates his/her strategy by imitating the strategy of a randomly selected neighbor **B** with probability [19, 20]

$$p_{fermi} = \frac{1}{1 + \exp(-\beta(f_B - f_A))} \quad (2)$$

where β is the intensity of selection (which regulates the randomness of the decision making process) and f_i is the fitness of individual i , which here is simply the accumulated payoff from all his interactions. Consequently, individuals imitate preferentially those peers that have a higher fitness, which may also be interpreted as a measure of social success.

Each *network configuration* was extracted from a time range of 2000 generations after an initial transient of 5 generations. Each evolution started from a random configuration

of traits. During the observation time window, and whenever the system was at the desired trait configuration j/Z , the associated *network configuration* was stored with probability $\sim 10^{-3}$, thus lowering the chances of storing two consecutive *network* configurations.

On both **HoSW** and **HeSW** networks we analyze the peer influence patterns for $\lambda = 0.01$ while on **EXP** and **BA** populations we considered a harsher social dilemma with $\lambda = 1.25$. These values of λ can be shown to encapsulate the regimes of interest for the networks topologies under consideration [21-23].

Once again we found that the critical network distance, n_c , tends to increase with increasing **APL** (white disks) — see [Figure SM.9](#) — increasing as well with decreasing $\langle k \rangle$ or increasing **CC**.

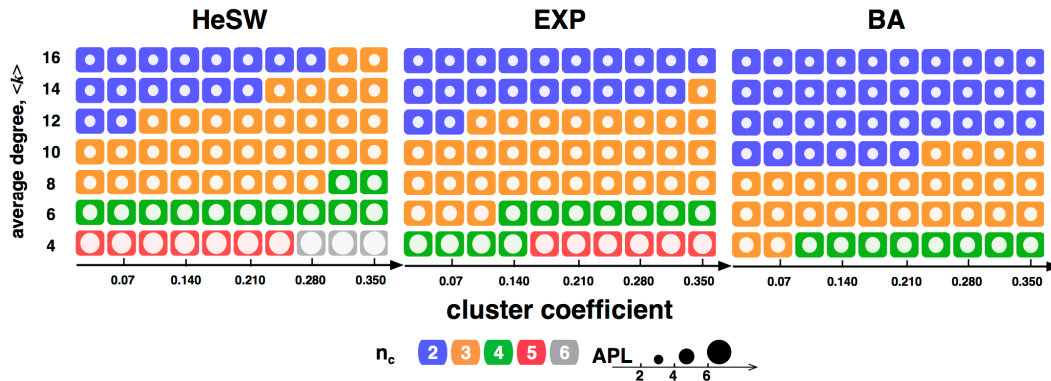


Figure SM.9 – Critical network distance on heterogeneous social networks with different average degrees and cluster coefficient. Populations have $Z = 10^3$ individuals, each colored cell denotes a different network and circles correspond to their *average path length (APL)*, which decreases with increasing *average degree* ($\langle k \rangle$) and decreasing cluster coefficient (**CC**).

We take into account the impact of each type of topology in the evolutionary dynamics of populations engaging in a **PD** by dividing our analysis in two different levels of λ : $\lambda=0.01$ ([Figure SM.10](#)) and $\lambda=0.25$ ([Figure SM.11](#)).

[Figures SM.10](#) and [SM.11](#) show the impact of population size (10^3 , 10^4 and 10^5) on the peer influence patterns generated from the **EGT** dynamics.

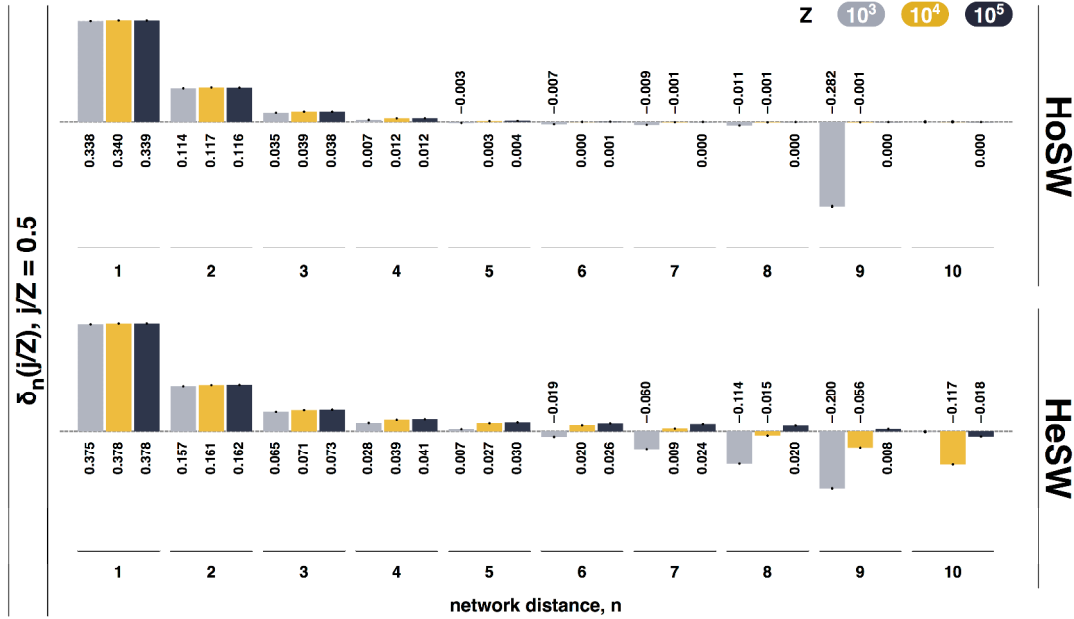


Figure SM.10 – Peer Influence patterns on Homogeneous (upper panel) and Heterogeneous (bottom panel) Small World populations with different sizes. We assume populations composed by $Z = 10^3$ (gray), 10^4 (yellow) and 10^5 (black) individuals and average degree $\langle k \rangle = 4$ and show how the correlation values vary with the network distance.

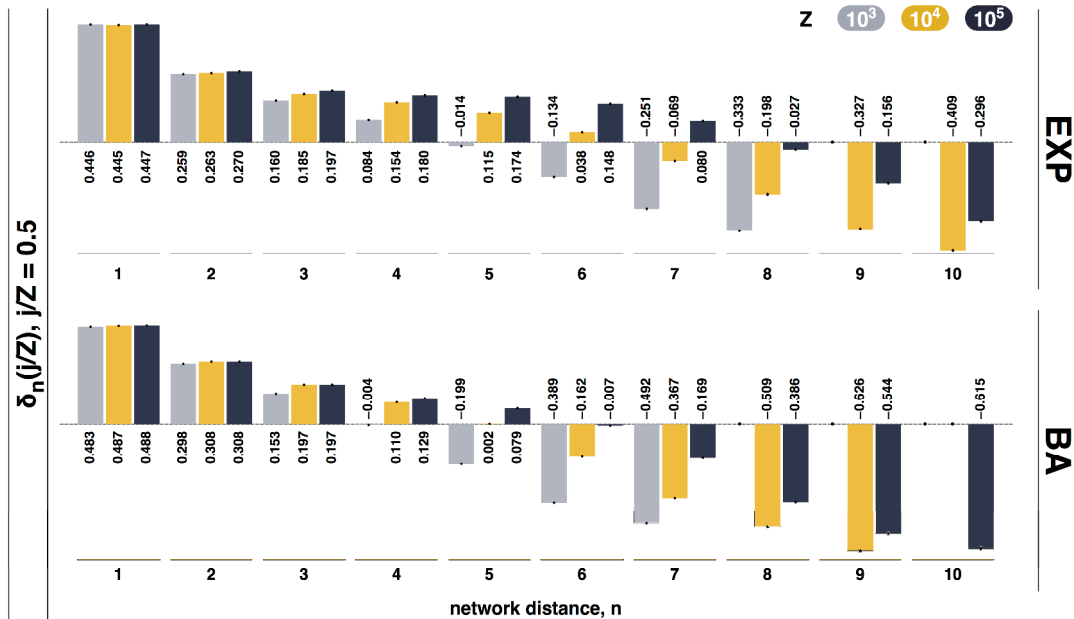


Figure SM.11 – Peer Influence patterns on Exponential (upper panel) and scale free (bottom panel) populations of different sizes. We assume populations composed by $Z = 10^3$ (gray), 10^4 (yellow) and 10^5 (black) individuals and average degree $\langle k \rangle = 4$ and show how the correlation values vary with the network distance.

On **HoSW** populations we observe a consistent pattern that tends to zero with increasing network distance n . On heterogeneous populations, such as **HeSW**, **EXP** and **BA**, we obtain qualitatively the same pattern found in previous dynamics: a monotonic decrease with increasing network distance.

6. References

- [1] F. C. Santos, F. L. Pinheiro, T. Lenaerts, and J. M. Pacheco, *J. Theor. Biol.* **299**, 88 (2012).
- [2] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, *Proc. Natl. Acad. Sci. U S A* **97**, 11149 (2000).
- [3] F. C. Santos, J. Rodrigues, and J. M. Pacheco, *Phys. Rev. E* **72**, 056128 (2005).
- [4] D. J. Watts, and S. H. Strogatz, *Nature* **393**, 440 (1998).
- [5] S. N. Dorogovtsev, and J. F. Mendes, *Adv. Phys.* **51**, 1079 (2002).
- [6] R. Albert, and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [7] M. E. Newman, *SIAM Rev.* **45**, 167 (2003).
- [8] S. Bansal, S. Khandelwal, and L. Meyers, *BMC Bioinformatics* **10**, 405 (2009).
- [9] J. H. Fowler, and N. A. Christakis, *Brit. Med. J.* **337** (2008).
- [10] S. Van Segbroeck, F. C. Santos, and J. M. Pacheco, *PLoS Comput. Biol.* **6**, e1000895 (2010).
- [11] R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. Lett.* **86**, 3200 (2001).
- [12] A. Barrat, M. Barthlémy, and A. Vespignani, *Dynamical processes on complex networks* (Cambridge University Press, 2008).
- [13] V. Sood, T. Antal, and S. Redner, *Phys. Rev. E* **77**, 041121 (2008).
- [14] T. Antal, S. Redner, and V. Sood, *Phys. Rev. Lett.* **96**, 188104 (2006).
- [15] T. M. Liggett, *Stochastic interacting systems: contact, voter and exclusion processes* (Springer, 1999), Vol. 324.
- [16] J. Maynard-Smith, *Evolution and the Theory of Games* (Springer, 1993).
- [17] J. Hofbauer, and K. Sigmund, *Evolutionary games and population dynamics* (Cambridge University Press, 1998).
- [18] K. Sigmund, *The calculus of selfishness* (Princeton University Press, 2010).
- [19] A. Traulsen, M. A. Nowak, and J. M. Pacheco, *Phys. Rev. E* **74**, 011909 (2006).
- [20] G. Szabó, and G. Fáth, *Phys. Rep.* **446**, 97 (2007).
- [21] F. L. Pinheiro, J. M. Pacheco, and F. C. Santos, *PLoS One* **7**, e32114 (2012).
- [22] F. C. Santos, J. M. Pacheco, and T. Lenaerts, *Proc. Natl. Acad. Sci. U S A* **103**, 3490 (2006).
- [23] F. L. Pinheiro, F. C. Santos, and J. M. Pacheco, *New J Phys* **14**, 073035 (2012).