

Dynamical origins of the community structure of an online multi-layer society

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2016 New J. Phys. 18 083045

(<http://iopscience.iop.org/1367-2630/18/8/083045>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 147.125.57.21

This content was downloaded on 02/12/2016 at 11:27

Please note that [terms and conditions apply](#).

You may also be interested in:

[Triadic closure dynamics drives scaling laws in social multiplex networks](#)

Peter Klimek and Stefan Thurner

[Irreducibility of multilayer network dynamics: the case of the voter model](#)

Marina Diakonova, Vincenzo Nicosia, Vito Latora et al.

[Comparative evaluation of community detection algorithms: a topological approach](#)

Günce Keziban Orman, Vincent Labatut and Hocine Cherifi

[Coupled dynamics of node and link states in complex networks: a model for language competition](#)

Adrián Carro, Raúl Toral and Maxi San Miguel

[An efficient community detection algorithm using greedy surprise maximization](#)

Yawen Jiang, Caiyan Jia and Jian Yu

[Enhancing community detection by using local structural information](#)

Ju Xiang, Ke Hu, Yan Zhang et al.

[Evaluating accuracy of community detection using the relative normalized mutual information](#)

Pan Zhang

[The effect of spatiality on multiplex networks](#)

Michael M. Danziger, Louis M. Shekhtman, Yehiel Berezin et al.

[Efficient exploration of multiplex networks](#)

Federico Battiston, Vincenzo Nicosia and Vito Latora



PAPER

Dynamical origins of the community structure of an online multi-layer society

OPEN ACCESS

RECEIVED

27 April 2016

REVISED

29 July 2016

ACCEPTED FOR PUBLICATION

8 August 2016

PUBLISHED

8 September 2016

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Peter Klimek^{1,5}, Marina Diakonova^{2,5}, Víctor M Eguiluz², Maxi San Miguel² and Stefan Thurner^{1,3,4}¹ Section for Science of Complex Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, A-1090, Austria² Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), E-07122 Palma de Mallorca, Spain³ IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria⁴ Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA⁵ These authors contributed equally to this work.E-mail: stefan.thurner@meduniwien.ac.at**Keywords:** multiplex network, voter model, social networks, community detection**Abstract**

Social structures emerge as a result of individuals managing a variety of different social relationships. Societies can be represented as highly structured dynamic multiplex networks. Here we study the dynamical origins of the specific community structures of a large-scale social multiplex network of a human society that interacts in a virtual world of a massive multiplayer online game. There we find substantial differences in the community structures of different social actions, represented by the various layers in the multiplex network. Community sizes distributions are either fat-tailed or appear to be centered around a size of 50 individuals. To understand these observations we propose a voter model that is built around the principle of triadic closure. It explicitly models the co-evolution of node- and link-dynamics across different layers of the multiplex network. Depending on link and node fluctuation probabilities, the model exhibits an anomalous shattered fragmentation transition, where one layer fragments from one large component into many small components. The observed community size distributions are in good agreement with the predicted fragmentation in the model. This suggests that several detailed features of the fragmentation in societies can be traced back to the triadic closure processes.

1. Introduction

Societies are organized in dynamical patterns that emerge from the social actions of individuals. These arrange an array of different types of social relationships (e.g. friendship, marriage, co-workers, ...) to form groups, organizations, or institutions of various sizes. Each type of social relation defines a social network of its own. The dynamics of the structure of these networks is coupled to the dynamics of the states of the nodes of the network, in what has been named as co-evolution network dynamics [1–4]. Additionally, these networks are not independent of each other but co-evolve with the other networks in the society, forming a multiplex network, i.e. a network where a set of nodes can be connected by links of more than one type [5, 6]. Societies can then be understood as a dynamic, co-evolving multiplex network. It has recently become clear that many properties of social networks can also be found in and generalized to social multiplex networks. For instance, a generic feature of social networks is that individuals tend to form communities, i.e. groups of nodes that share more links with each other than with nodes outside of the community [7, 8]. Social multiplex networks typically exhibit this community structure not only within, but also *across* different layers [9–11], i.e. two nodes that are in the same community in one of the layers have the tendency to belong to the same community in another layer. The topological structures of individual network layers can vary dramatically, depending on the type of the corresponding social interactions [12, 13]. For example, it has been shown that network layers corresponding to cooperative behavior can be characterized by high clustering, high reciprocity, and high link overlap, whereas layers encoding aggressive behavior exhibit pronounced power-law degree distributions [14, 15]. Consequently

one of the big challenges in the field is currently to devise models that are capable of explaining such observed heterogeneities between individual layers. It has been shown, for instance, that many characteristics of the individual multiplex layers, such as their degree distributions, clustering coefficients, or the probabilities for nodes to acquire new links, can be understood from the assumption that the link dynamics in networks is driven by the process of triadic closure (i.e. the tendency that nodes with common neighbors will establish links between themselves) together with a finite lifetime of links (i.e. the typical probability at which they are added to and removed from the network) [16–20]. Triadic closure has indeed been long known to play a central role in social link formation that is deeply rooted in human psychology [21, 22]. Similarly, the organization of societies into community structure follows such psychological principles, too. This has been observed in the ubiquitous, hierarchically nested organization of communities of different sizes [23, 24] or Dunbar’s number [25], a hypothesized upper cognitive limit to the number of people with whom humans can share stable social relationships.

In this work we investigate the dynamical origins of the community structure of societies in multiplex networks. We study how different layers in the multiplex network influence each other and investigate the resulting consequences of these interactions for the community structures in the individual layers. Community structure will be simply characterized by the community size distributions in the different network layers, i.e. the probability for an individual (node) to be part of a community of a given size in a particular layer. As a data set we use the comprehensive dynamic social multiplex network of the *Pardus* society [14, 15, 18, 26–28]. This is a virtual society of more than 380 000 players with different social and economic interactions taking place in the open-ended massive multiplayer online game *Pardus*. We find a substantial amount of heterogeneity in the community size distributions across the different layers.

We address whether the empirical observations can be understood as a possible manifestation of the so-called fragmentation transition. Fragmentation is the phenomenon in which a network might undergo a transition from a state where almost all nodes belong to a single giant component to a *fragmented* state in which the network breaks into several components. The fragmentation transition is a phenomenon generically associated with co-evolution network dynamics [29–34]. Here we consider the voter model (VM) [35, 36] as a paradigmatic example. In the co-evolutionary VM (CVM) [31] a node can either change its internal state to that of its neighbor or re-wires one of its existing links towards a node that has the same internal state. The CVM has a fragmentation transition for a threshold value of the link rewiring rate. In an extension of the CVM to multiplex networks (MCVM) some, but not all, nodes are interconnected across layers. Such nodes then have the same internal states in both layers [33]. This MCVM has an anomalous transition called *shattered fragmentation* in which one layer assumes a fragmented state whose topological properties would be different under the same model parameters in a single layer CVM. It has been shown that the driving force behind this transition is the asymmetry between re-wiring probabilities in different network layers, i.e. the individual lifetime of links in each of the layers [33].

To reconcile the dynamics of the CVM with the empirical dominance of triadic closure, where a substantial amount of newly created links in social networks connect nodes that already share common neighbors [15, 18, 19, 21, 22], we propose a novel type of CVM in which the re-wiring step is carried out by a triadic closure process. We study the dynamics of this model on, both, single-layer (TCVM) and multiplex networks (MTCVM). We show that the triadic closure models display a novel fragmentation transition that is again different from the shattered fragmentation found in the MCVM. We investigate whether the fragmentation behavior of the MTCVM is compatible with the community structure observed in *Pardus* multiplex network data. To understand the extent to which the mesoscopic structure of the model multiplex network is compatible with the data, we compare results for pairwise inter-layer similarities of the detected communities.

We present the *Pardus* data and describe our methods in section 2. The TCVM and its generalization to multiplex networks (MTCVM) is introduced in section 3. The community structure predicted by the MTCVM is compared to results from the data in section 4 and conclusions are drawn in section 5.

2. Data and methods

2.1. Data

The *Pardus* dataset contains all actions of more than 380 000 players in a massive multiplayer online game. The players interact in a virtual, open-ended game Universe to connect with other players to achieve self-posed goals, such as accumulating wealth and influence. Players can engage in three different types of cooperative interactions. They can establish mutual friendship links, exchange private, one-to-one messages and trade with each other in the game. The data can be represented as a dynamic multiplex network $M^\alpha(t)$, where the index α labels the adjacency matrix of the network given by interactions of type α at time t . The multiplex network $M^\alpha(t)$ is constructed for each month (30 days) over one year of data, from Sep 2007 to Sep 2008. Two players are

Table 1. Overview of characteristics of the Pardus multiplex network layers. For each layer α we show the values of the re-wiring probability p_α and the average number of links, L^α . The re-wiring probability p_α is orders of magnitude smaller for the friendship network, compared to communication and trade. Each layer has the same average number of nodes, $N = 3.1(2) \times 10^3$.

α	Friendship	Trade	Communication
p_α	0.004(1)	0.27(1)	0.35(2)
L^α	$1.45(2) \times 10^4$	$5.57(2) \times 10^4$	$1.90(3) \times 10^4$

linked in a corresponding multiplex layer if they had a friendship link ($\alpha = \text{friend}$), traded with each other ($\alpha = \text{trade}$), or exchanged a private message ($\alpha = \text{communication}$) within a given month. For a particular t we include all players that have at least one link in each of the multiplex layers. Nodes that are isolated in at least one layer at a given time are pruned. As a consequence of this removal other nodes might become isolated, so the pruning has to be iterated until there are no isolated nodes. For more information on the topology and structure of the Pardus multiplex network see [14, 15]. Per construction, each layer has the same average number of non-isolated nodes (i.e. players that are active in that particular time interval), $N = 3.1(2) \times 10^3$. Numbers in brackets denote standard deviations. Here and in the following, we drop the time dependence of a variable whenever we refer to its average value over all time intervals. Table 1 shows results for the average number of links in each layer, L^α . The trade network shows the highest link density, the friendship and communication network have similar numbers of links.

It has been shown that for the given time-span the three network layers α are in a stationary state in the sense that links are added and removed with comparable rates. These probabilities are orders of magnitude larger than the probabilities at which nodes are added or removed, see [18]. The dynamics within the network layers is therefore dominated by re-wiring processes. The re-wiring probability p_α is defined as the average value of the probabilities that a link will be added or removed, that is, re-wired, in layer α . Let $\Delta I_\alpha^{+/-}(t)$ denote the number of links that are added/removed at t . The re-wiring probability for t , $p_\alpha(t)$, is then $p_\alpha(t) = \frac{\Delta I_\alpha^+(t) + \Delta I_\alpha^-(t)}{2L^\alpha(t)}$. Table 1 shows results for the average re-wiring probabilities p_α in the three layers. p_α in the friendship network is orders of magnitude smaller than in the communication and trade networks. This can be understood by the difference in processes that govern the interactions in these layers. In the friendship network a link persists after it has been formed until the link is removed or one of the players leaves the game. In the message and communication network, on the other hand, a link is only formed between two players if at least one interaction took place within the considered time interval. This results in a substantially lower turnover of links in the friendship network than in the other layers. We will therefore refer to the friendship layer as being *slow* and to the trade and message layers as being *fast* in terms of the average time between two consecutive re-wiring events in the given layer. Note that although friendship links have the longest survival time (i.e. lowest turnover p_α), the friendship network has also the smallest number of links L^α .

2.2. Community detection

From the network layers $M^\alpha(t)$, for data and model, we identify the communities $C^\alpha(t) = \{C^\alpha(1, t), C^\alpha(2, t), \dots, C^\alpha(N_c^\alpha, t)\}$, where the i th community $C^\alpha(i, t)$ is the set with a number of $n_i^\alpha(t)$ nodes within community i at time t . We use four different community detection algorithms together with a clean-up procedure that allows us to detect the *absence* of community structure as well as *homeless* nodes, which do not belong to any community [37]. Those four algorithms are OSLOM [37], Infomap [38], Copra [39], and the Louvain algorithm [40]. OSLOM's clean-up procedure is used with a coverage parameter of 1 and a standard significance threshold of $p < 0.1$ [37] to ensure that we retrieve only statistically significant communities and avoid the so-called *resolution limit* problem (failure of the detection of small communities) [41]. Results for the communities in the Pardus data are discussed in section 4.1.

2.3. Components in the model

To understand the fragmentation behavior of the MTCVM we describe the organization of the model-networks into *components* by the following observables. $N_c^\alpha(t)$ is the number of components in network layer α at time t . We will refer to the time average of $N_c^\alpha(t)$ over T consecutive time-steps by dropping the dependence on t , i.e. $N_c^\alpha \equiv \frac{1}{T} \sum_t N_c^\alpha(t)$. The size of the k -largest component at t is denoted $S_k^\alpha(t)$. The time average of the k -largest component size, S_k^α , is again denoted by dropping the dependence on t .

2.4. Normalized mutual information

The similarity of the sets of links in two layers, $M^\alpha(t)$ and $M^\beta(t)$, is measured by the normalized mutual information, $\text{NMI}(\alpha, \beta; t)$ [37, 42, 43], given as follows. Let $H(\alpha; t)$ be the entropy of the set of communities $C^\alpha(t)$, and $H(\beta|\alpha; t)$ be the conditional entropy of communities $C^\beta(t)$, conditioned on $C^\alpha(t)$, as defined e.g. in [42, 43]. The mutual information between the sets of communities in $M^\alpha(t)$ and $M^\beta(t)$, $I(\alpha; \beta; t)$, is then given by

$$I(\alpha; \beta; t) = \frac{1}{2}(H(\alpha; t) + H(\beta; t) - H(\beta|\alpha; t) - H(\alpha|\beta; t)). \quad (1)$$

Typically, $I(\alpha; \beta; t)$ is normalized such that it takes on a value of zero if the overlap between the communities in layers α and β is completely random, and a value of one if the communities are identical. We use a normalization that has been shown to be particularly suited for situations, where $C^\alpha(t)$ and $C^\beta(t)$ have very dissimilar numbers of communities [43]. The normalized mutual information, $\text{NMI}(\alpha, \beta; t)$, is then

$$\text{NMI}(\alpha, \beta; t) = \frac{I(\alpha; \beta; t)}{\max[H(\alpha; t), H(\beta; t)]}. \quad (2)$$

As before, we refer to the time average of $\text{NMI}(\alpha, \beta; t)$ by dropping time dependence.

3. Model

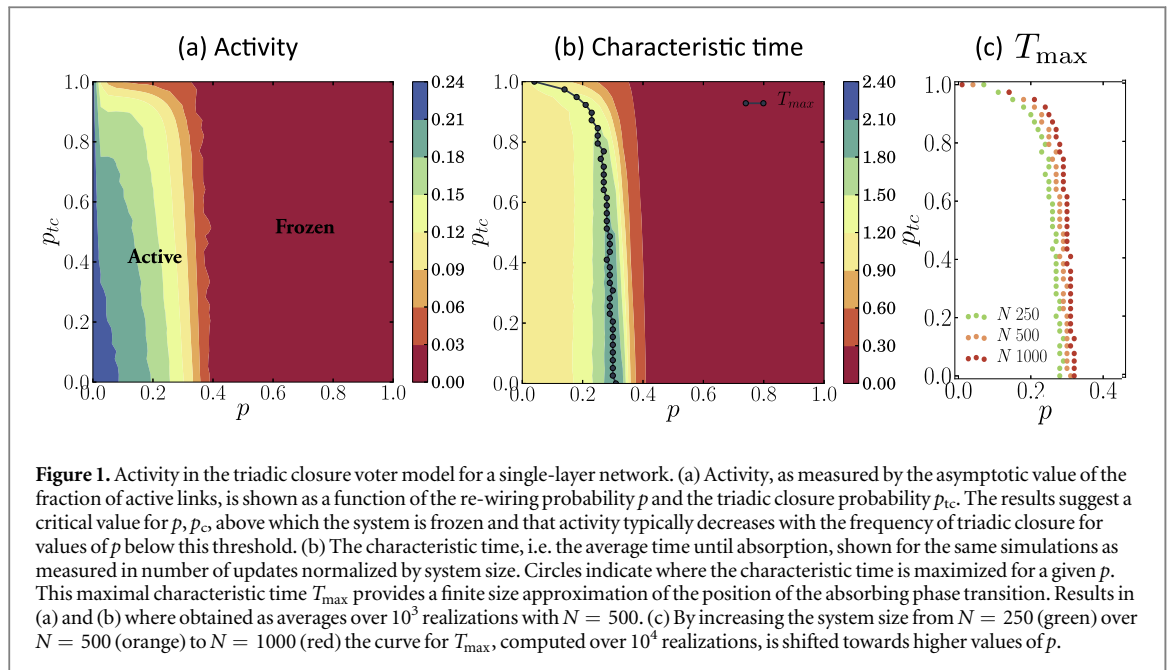
3.1. Triadic closure VM on a single-layer network

The standard binary-state CVM on a single network as introduced in [31] is given as follows. Each node, i , is described as a time-dependent, binary, internal state $\sigma_i(t) \in \{0, 1\}$ subject to the following update rule. (i) Pick node i and one of its neighbors, j , at random. If the internal states of these nodes differ, $\sigma_i(t) \neq \sigma_j(t)$, then (ii-a) with probability p the link between i and j is removed and a link between i and a different node k is formed. k is randomly chosen from the set of all nodes disconnected to, but in the same internal state as node i , $\sigma_i(t) = \sigma_k(t)$. If no such node k exists, the link is not re-wired. Otherwise, (ii-b) the state of node i is changed to $\sigma_j(t)$. This model has one parameter, the re-wiring probability p , which defines the preference of a node to re-wire its link or change its internal state, if the node has a neighbor with a different internal state. Thus, p sets the ratio of time scales between the dynamics of the network and the dynamics of the state of the nodes.

We now introduce the triadic closure VM (TCVM) on a single network. The model is motivated by the empirical fact that links that connect nodes that share common neighbors are more likely to form than links that do not connect such nodes, i.e. the process of triadic closure [18, 19, 21, 22]. The re-wiring step (ii-a) of the CVM is now replaced by the following update rule. With the triadic closure probability, p_{tc} , the new link is made between node i and a different node l that is randomly chosen from the set of all nodes that share at least one common neighbor with i (but there is no connection between i and l , yet). If no such node l exists, no re-wiring takes place. Otherwise, with probability $1 - p_{\text{tc}}$, we follow the re-wiring rule from step (ii-a) of the CVM.

In the simulations the network is initialized as a random regular network with size N and degree $\mu = 4$ for each node, which results in a network that is initially connected. In [33, 34] shattered fragmentation was demonstrated on networks with these parameters, and we use them here to facilitate comparisons to earlier works. These prior works suggest that changing the initial intra-layer connectivity does not lead to major qualitative differences in the phase diagram of the model and that an average degree of $\mu = 4$ is representative of a wide range of values for μ . Moreover, we choose networks with homogeneous degree distributions to avoid having to consider an additional parameter for the (anti-)correlation of intra-layer degrees when we later extend the model to multiplex networks. The initial distribution of internal states is random, so that each node has the same probability $1/2$ to be in one of the two possible states. We call a link *active* if it connects two nodes with different internal states. If there are no active links, the system is in an absorbing configuration. To measure whether large systems ($N \rightarrow \infty$) reach an absorbing state, we examine system activity, which is given by the asymptotic value of the averaged interface density [36, 44]. The interface density measures the fraction of active links and is zero in absorbing states. The CVM with random re-wirings (i.e., $p_{\text{tc}} = 0$) on a finite single-layer network always reaches an absorbing state. This is also the case for partial triadic closure rewiring $p_{\text{tc}} > 0$. The extreme case of pure triadic closure rewiring given by $p = 1$ and $p_{\text{tc}} = 1$ allows for the existence of network traps, where active links between nodes in different states remain. This happens exactly when an active link between i and j is also their *only* active link. An isolated node pair in different internal states is one such example. Clearly as long as $p_{\text{tc}} < 1$ or $p < 1$ this trap is avoided as either one of the nodes re-wires to yet another node, or eventually one of the nodes changes its state. Another consequence of re-wirings purely by triadic closure ($p_{\text{tc}} = 1$) is that once a node is isolated, it cannot be brought back into a component, hence the number of isolated components, N_c , is non-decreasing over time in this case.

Systems with finite N always reach absorbing states through finite-size fluctuations. The asymptotic value of the average of the interface density is a good proxy for the behavior of large systems [31]. Such systems will reach

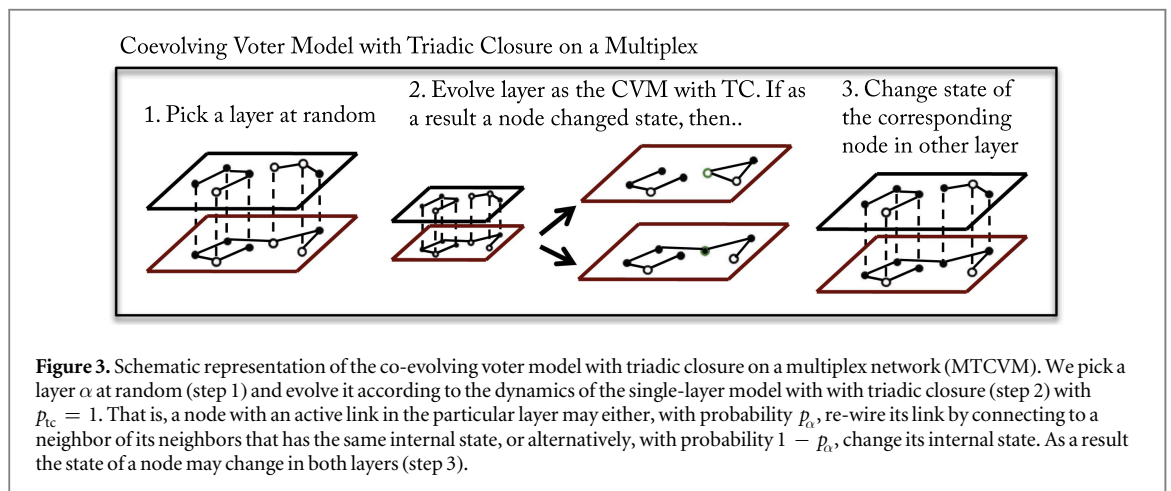
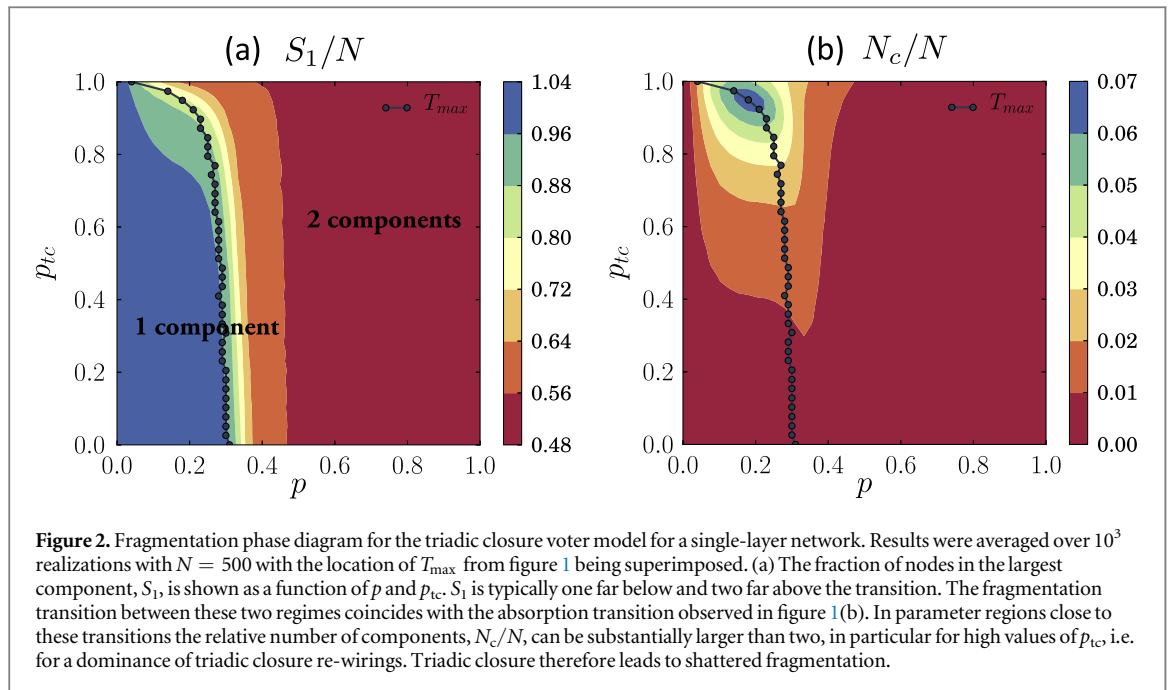


absorbing states (freeze) if the activity is zero. We find that large systems always reach a frozen, absorbing configuration for values of the re-wiring probability p that are above a critical threshold, p_c . Below this threshold we find that systems stay active, and that their activity decreases with increasing probability of triadic closure, p_{tc} . In the limit of $p_{tc} = 0$, corresponding to the CVM [31] with random re-wirings only, activity decreases with re-wiring and falls to zero around a critical re-wiring probability p_c , which is defined in the limit of infinite systems. This critical re-wiring probability characterizes an absorbing phase transition and is accompanied by an increase of the time it takes for the system to reach an absorbing state. The average time-to-absorption gives the characteristic time of the system, which is shown in figure 1(b), measured in number of updates normalized by system size. For a given value of p_{tc} we associate its peak value $T_{max}(p_{tc})$ with the finite-size approximation of the position of the absorbing transition, $p_c(N)$. Figure 1(b) shows that increasing the frequency of triadic closure not only decreases activity, but also the critical re-wiring probability p_c . In other words, triadic closure causes systems to freeze for even lower re-wiring probabilities p . Figure 1(c) shows that larger values of the system size N typically lead to an increase in p_c . Furthermore, figure 1 suggests that the absorbing transition does not exist for $p_{tc} = 1$; large systems that re-wire *only* with triadic closure do not sustain a constant level of activity and will always reach an absorbing state in finite time.

The absorbing and the fragmentation transition coincide (figure 2). For ($p < p_c$), systems in the thermodynamic limit remain active, whereas systems with finite N reach an absorbing state that can be characterized by one giant component with all nodes in the same state. For ($p > p_c$), independent of system size, systems will reach an absorbing state characterized by having two components that correspond to the two initially present states. The presence of triadic closure changes the fragmentation behavior of the system in a peculiar way that is most obvious for large values of p_{tc} that are close to p_c . In this region of parameters the decrease in the size of the largest component is not compensated only by the growth of the second largest component, but also by an increase in the number of components, N_c . Most of these components consist only of a small number of nodes. This is the hallmark of a shattered fragmentation; topologies with at least one giant components and a multiplicity of components of smaller size. Shattered fragmentation was first observed in the CVM in multiplex networks where not all nodes are present in each layer (MCVM) [33] and for the CVM on a single-layer network with noise that targets a fixed subpopulation of nodes [34]. Here, we find the same phenomenology for re-wiring processes that are dominated by triadic closure processes.

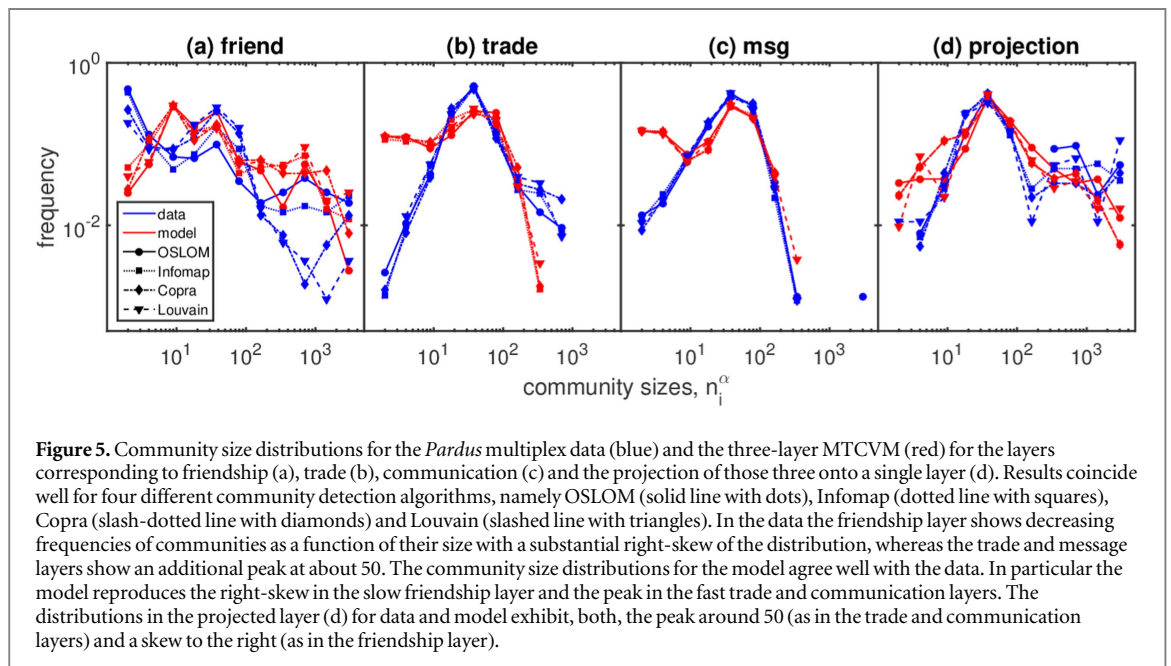
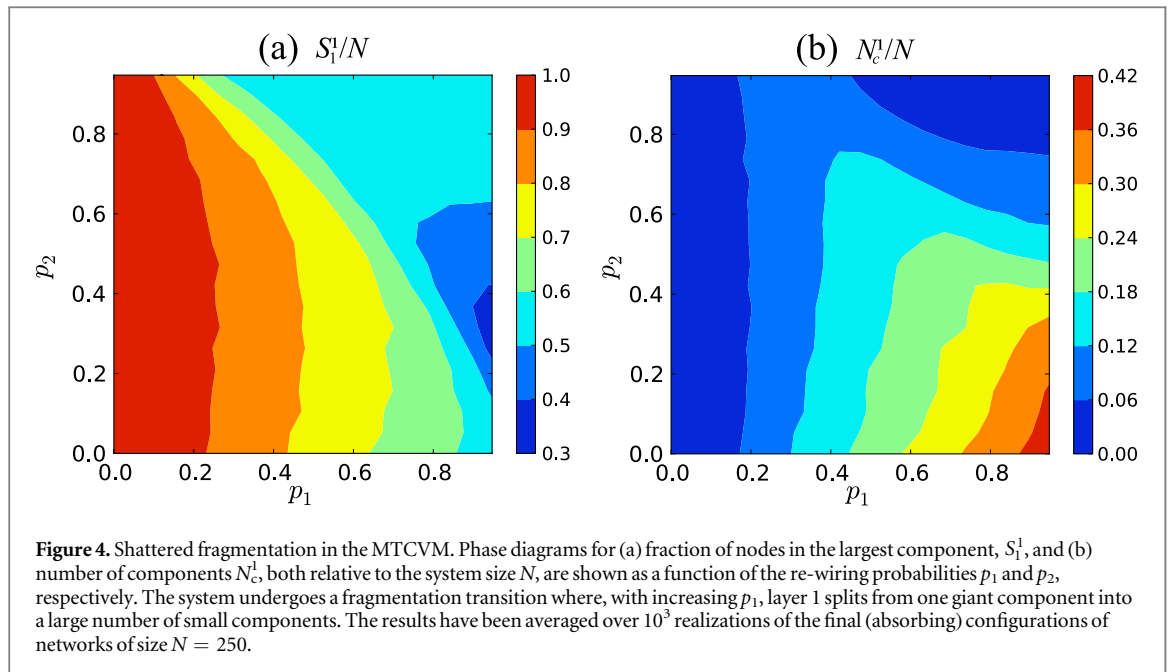
3.2. The TCVM on multiplex networks (MTCVM)

We now introduce the MTCVM that is composed of two layers that are fully pair-wise interconnected (each nodes has its counterpart in another layer). Each of the layers has its own re-wiring probability, p_1 and p_2 , respectively. The multiplex network evolves by, first, picking a layer α , $\alpha \in \{1, 2\}$, randomly and, secondly, by evolving that layer by using the triadic closure re-wiring rule, see figure 3. This introduces a co-evolution of the two layers because the formation of new links depends on the internal state of nodes which is the same in both layers for each of the nodes. We study the behavior of this model for $p_{tc} = 1$, i.e. for the case where the re-wiring is done by triadic closure, that is without random re-wirings.



The MTCVM displays a novel fragmentation transition that is visible in the phase diagrams of the observables S_1^α (size of largest component) and N_c^α (number of components). Figure 4 shows the phase diagrams of (a) S_1^1 and (b) N_c^1 in layer 1 of the MTCVM as a function of the re-wiring probabilities p_1 and p_2 . Consider the case where layer 1 is connected to a static layer with $p_2 = 0$. With increasing p_1 the largest component of layer 1 shrinks to a value around 0.5, similar to the single-layer case. N_c , on the other hand, increases by increasing p_1 in the same way as S_1 decreases. This means that with increasing p_1 nodes leave the largest component and form small or isolated components of their own (and not a second, large component), i.e. N_c increases. This is again the process of shattered fragmentation [33]. From the phase diagrams in figure 4 it follows (by exchanging the labels of layers 1 and 2) that in the static layer 2 the nodes remain in one giant component as layer 1 undergoes this shattered fragmentation. The number of components in the final fragmented state is maximized for the layer with higher p as the re-wiring rate in the other layer goes to zero (bottom right hand corner of figure 4(a)). The slower layer will not fragment but freeze in one giant component. If one increases its re-wiring rate, it will freeze not in one, but in two giant components that correspond to the two internal states of nodes existing at initialization. The number of small components in the faster layer will decrease at the expense of building up the second largest component. Finally, when both layers have re-wiring rates close to on, they will freeze in two giant components only. The relative size of the second largest component (not shown) starts to grow in the second quarter of the diagram, peaking in the top-right corner at 0.5.

Note that this fragmentation transition in the MTCVM differs in some respects from the fragmentation transition observed in the MCVM with random re-wirings. This model can be recovered by setting $p_{tc} = 0$ in



both layers and by introducing an additional parameter, the multiplexity q [33]. That is, both layers contain N nodes, but only a subset of qN nodes exists in both layers. The internal states of the remaining $(1 - q)N$ nodes depend only on the dynamics in their own layer. The multi-layer CVM with random re-wirings also displays a shattered fragmentation transition that is, however, only encountered below a critical value of q , $q < q_c < 1$. Under triadic closure, partial multiplexing (i.e. $q < 1$) is no longer required to see shattered fragmentation.

4. Empirical results

4.1. Community structure in the *Pardus* multiplex network

The time-averaged distribution functions for the community sizes in the three network layers are shown in figures 5(a)–(c). There is a clear discrepancy between the distribution observed in the *slow* friendship layer (a) when compared to the trade (b) and message (c) networks, which are characterized by substantially larger re-wiring probabilities. For the friendship layer the frequencies of community sizes resemble a monotonously decreasing function in size that develops a fat tail. In contrast, there exist distinct peaks in the distributions of

community sizes in the trade and message layers, which are also both *fast* layers. These peaks are centered around a community size of 50.

4.2. The MTCVM versus *Pardus* data

So far we have discussed the behavior of the MTCVM on multiplex networks that consist of two layers. The *Pardus* data, however, consists of three layers. The extension of the MTCVM to three layers is straightforward. In the first step of the algorithm one randomly chooses one of the layers for the update. The only change in the three-layer-case is that one now picks one out of *three* layers with equal probability. The resulting model has now three parameters, namely the re-wiring probabilities of each layer. A thorough exploration of the corresponding enlarged phase space is beyond the scope of the current work. To calibrate the model to the data we assume that the model internal states, $\sigma_i(t)$, encode a hidden propensity of the players to interact with each other. That is, upon meeting two players i and j will be more likely to establish a link with each other if they have the same internal states. We further assume that as initial condition a link is active with equal probability in each of the layers. A re-wiring of this link occurs then with the measured probability p_α in layer α in, both, model and data. Note that this definition does not fix all the time scales of the model, for which we would also need to take the probability to encounter an active link in the layers into account. Two layers in this model, the ‘fast’ layers of trade and communication, have very similar values of p_α that are orders of magnitude larger than the ‘slow’ friendship layer. It is therefore reasonable to assume that for this particular realization of a three-layer MTCVM we find a similar behavior as for a two-layer MTCVM that consists of a fast and a slow layer. Our simulations confirm that the three-layer MTCVM indeed displays the same shattered fragmentation behavior that was observed in the corresponding two-layer model.

The community size distributions obtained from the model agree well with the data. In figure 5 we show the frequency of community sizes observed in ten realizations using $N = 3100$ and the measured re-wiring probabilities p_α as measured in the data. Results are shown for the (a) friendship (b) trade, and (c) communication layer and for four community detection algorithms. The different algorithms give very similar results. The frequencies of community sizes in the friendship layer in the model decrease as a function of their size in strong resemblance to results observed in the data. For the trade and communication layers, respectively, we also observe a clear peak in the frequencies of community sizes that coincides with the peaks observed in the data around a community size of 50. Note that here we compare communities in the data to communities in the model (and not to components). Clearly, the fragmentation behavior into components in the model is closely related to its community structure. Given the re-wiring probabilities p_α in the data, we would expect from the MTCVM that the friendship layer ($p^{\text{friend}} = 0.004(1)$) shows one large component whereas the trade ($p^{\text{trade}} = 0.27(1)$) and communication ($p^{\text{comm}} = 0.35(2)$) layers display fragmentation. Indeed, we observe for the *communities* in, both, data and model for the friendship layer a heavily right-skewed, if not fat-tailed, size distribution with a small number of very large communities, whereas the trade and message layers fragment into a large number of smaller communities with a peak centered at around 50. Data and model show the same behavior in terms of fragmentation into communities of various sizes. Figure 5(d) shows the community size distributions for networks that are obtained by projecting the three layers onto a single network. That is, the projection networks for data or model contain a link between two nodes if they are connected in *any* of the three data or model layers, respectively. Again, the observed distributions of community sizes are similar in data and model. They show both features that characterize the distributions in the individual layers, i.e. a peak around sizes of 50 together with a substantial right-skew of the distributions.

A statistically sound and robust measure to quantify the similarity of two different sets of communities is the normalized mutual information [37, 42, 43, 45]. We therefore investigate the similarities between communities in the data and model by comparing the normalized mutual information, $\text{NMI}(\alpha, \beta)$, between each pair of layers in the data and model multiplex networks, respectively. When averaged over the results for each community detection algorithm and over all timesteps or realizations for data and model, respectively, we find values of $\text{NMI}(\alpha, \beta)$ that are about one order of magnitude larger in the data as in the model. In particular we find for friendship–communication for the data $\text{NMI}(\alpha, \beta) = 0.04(5)$ and $\text{NMI}(\alpha, \beta) = 0.004(1)$ in the model. For friendship–trade we find $\text{NMI}(\alpha, \beta) = 0.03(3)$ in the data and $\text{NMI}(\alpha, \beta) = 0.003(1)$ in the model. Finally, for communication–trade in the data we have $\text{NMI}(\alpha, \beta) = 0.15(5)$ and $\text{NMI}(\alpha, \beta) = 0.005(1)$ in the model. First, note that the values in the data are clearly larger than the model results. This can be understood by the fact that the MTCVM does not contain any mechanism that explicitly increases the similarity of layers, such as by copying one link from one layer to the other. However, if we restrict the analysis to only the largest community in each layer and discard all other communities, we find that the resulting NMI values in data and model become of similar magnitude (between 0.69 and 0.86). This means that the tendency of nodes to belong to the largest community in two different layers in the data is well approximated by the assumption of a shared, internal state in the model. The finer structure of communities with a smaller number of nodes, however, is not captured in the model. Intriguingly we find that in, both, data and model the

communities in the fast layers (trade and communication) are more similar to each other than the communities of a fast and a slow layer (i.e. friendship–trade and friendship–communication, respectively). That is, we find higher NMI values between the communication–trade layers than between the friendship–communication layers in both data ($p_{\text{WSR}} < 10^{-7}$ as the p -value from a one-sided, paired Wilcoxon signed rank test) and model ($p_{\text{WSR}} < 10^{-6}$). The same observation holds for the NMI values of the communication-trade layers when compared to the friendship-trade results in data ($p_{\text{WSR}} < 10^{-7}$) and model ($p_{\text{WSR}} < 10^{-6}$). That is, even though the inter-layer similarity properties of the communities do not agree between data and model in absolute values as measured by the NMI, their results relative to other pairs of layers in the same multiplex network *do* actually compare well. This is a remarkable result in the absence of an explicit ‘link-copying’ mechanism between the individual layers.

5. Discussion

We investigated the dynamical origins of the community structure of societies represented as dynamical multiplex networks. In empirical data from the large-scale online game society *Pardus* we observed substantial differences in the community structures of individual layers of this multiplex network. While one layer is characterized by a small number of large communities and a decreasing distribution of community sizes, in the other layers we find a peak of communities of intermediate size around 50. We found that the time scales on which the link dynamics takes place in the various network layers differ by several orders of magnitude. Remarkably, we find that the monotonously decreasing distribution of community sizes is found in a layer with a very small re-wiring probability, whereas layers with the centered distribution are characterized by probabilities that are orders of magnitude higher.

To understand these empirical findings we proposed a generalization of the CVM on multiplex networks which incorporates the process of triadic closure, the MTCVM. This process has been shown to be crucial in modeling the structure formation of individual layers in the *Pardus* society [18]. We studied the phase diagram of the new model and found that it exhibits an anomalous fragmentation transition for multiplex networks that makes the model interesting in its own right. This transition is characterized by a break-up of the largest component of a network layer into a large number of small components. Intriguingly, a key role in the model turns out to be the heterogeneity in time-scales on which the link re-wiring dynamics takes place in the individual layers. When the model is calibrated to mimic the *Pardus* data on a three-layer multiplex network, community size distributions are perfectly compatible with those found in the data.

In particular the model confirms that *slow* layers in terms of the time-span between two re-wiring events show a monotonously decreasing distribution of community sizes, whereas the *fast* layers display an additional peak around community sizes of 50. This means that the empirical community structure of the *Pardus* virtual society indeed resembles the fragmentation behavior predicted by the MTCVM. Note that for these results the model layers only differ in their re-wiring probabilities (but not, for instance, in their degree), so that these results can only be attributed to the multiplex interaction of dynamics on different time-scales. The focus of the current work was to understand how the community structure within a given layer arises from underlying microscopic multi-level interactions, wherein we depart from other interesting approaches that aim at identifying community structure *across* different layers in the multiplex network [9, 11]. We further confirmed that the communities in the fast layers of the multiplex networks are more similar to each other than they are similar to the slow layer in data and model as well. These results suggest that the dynamical origin of the community structure of societies can be understood through the interplay of triadic closure processes taking place on different time-scales, which manifests itself in the phenomenon of shattered fragmentation. Whether these results hold for empirical data in real-world societies remains to be seen.

Acknowledgments

We acknowledge financial support from FP7 projects LASAGNE (to MD, VE, MSM, and ST), agreement no. 318132, and MULTIPLEX, agreement no. 317532 (to PK and ST). MSM also acknowledges financial support from FEDER (EU) and MINECO (Spain) under Grant ESOTECOS (FIS2015-63628-C2-2-R).

Competing interests

The authors declare that they have no competing interests.

Author contributions MD, MSM, PK, ST and VME designed the research, MD and PK analyzed the data, developed the models and wrote the paper, MSM, ST and VME conceived and co-wrote the paper.

References

- [1] Zimmermann M G, Eguíluz V M and San Miguel M 2001 *Economics with Heterogeneous Interacting Agents (Lecture Notes in Economics and Mathematical Series vol 503)* (Berlin: Springer) pp 73–86
- [2] Zimmermann M G, Eguíluz V M and San Miguel M 2004 *M. Phys. Rev. E* **69** 065102
- [3] Gross T and Blasius B 2008 *Proc. R. Soc. Interface* **5** 259–71
- [4] Vazquez F 2013 *Dynamics On and Of Complex Networks (Modeling and Simulation in Science, Engineering and Technology vol 2)* ed A Mukherjee et al (New York: Springer) pp 89–107
- [5] Boccaletti S, Bianconi G, Criado R, del Genio C, Gómez-Gardeñes J, Sendiña-Nadal I, Wang Z and Zanin M 2014 *Phys. Rep.* **544** 1–22
- [6] Kivela M, Arenas A, Barthelemy M, Gleeson J P, Moreno Y and Porter M A 2014 *J. Complex Netw.* **3** 203–71
- [7] Girvan M and Newman M E J 2002 *Proc. Natl Acad. Sci. USA* **99** 7821–6
- [8] Fortunato S 2010 *Phys. Rep.* **486** 75–174
- [9] Mucha P J, Richardson T, Macon K, Porter M A and Onnela J-P 2010 *Science* **328** 876–8
- [10] Murase Y, Török J, Jo H-H, Kaski K and Kertész J 2014 *Phys. Rev. E* **90** 052810
- [11] De Domenico M, Lancichinetti A, Arenas A and Rosvall M 2015 *Phys. Rev. X* **5** 011027
- [12] Nicosia V, Bianconi G, Latora V and Barthelemy M 2013 *Phys. Rev. Lett.* **111** 058701
- [13] Kim Y Y and Goh K-I 2013 *Phys. Rev. Lett.* **111** 058702
- [14] Szell M, Lambiotte R and Thurner S 2010 *Proc. Natl Acad. Sci. USA* **107** 13636–41
- [15] Szell M and Thurner S 2010 *Soc. Netw.* **39** 313–29
- [16] Davidsen J, Ebel H and Bornholdt S 2002 *Phys. Rev. Lett.* **88** 128701
- [17] Holme P and Kim B J 2002 *Phys. Rev. E* **65** 026107
- [18] Klimek P and Thurner S 2013 *New J. Phys.* **15** 063008
- [19] Bianconi G, Darst R K, Iacovacci J and Fortunato S 2014 *Phys. Rev. E* **90** 042806
- [20] Battiston F, Iacovacci J, Nicosia V, Bianconi G and Latora V 2016 *PLoS One* **11** e0147451
- [21] Simmel G 1908 *Soziologie: Untersuchungen Über Die Formen der Vergesellschaftung* (Berlin: Duncker & Humblot)
- [22] Heider F 1946 *J. Psychol.* **21** 107–12
- [23] Hill R A and Dunbar R I M 2003 *Human Nat.* **14** 53–72
- [24] Fuchs B, Sornette D and Thurner S 2014 *Sci. Rep.* **4** 6526
- [25] Dunbar R I M 1993 *Behavi. Brain Sci.* **16** 681–735
- [26] Szell M, Sinatra R, Petri G, Thurner S and Latora V 2012 *Sci. Rep.* **2** 457
- [27] Thurner S, Szell M and Sinatra R 2012 *PLoS One* **7** e29796
- [28] Szell M and Thurner S 2013 *Sci. Rep.* **3** 1214
- [29] Vazquez F, González-Avella J C, Eguíluz V M and San Miguel M 2007 *Phys. Rev. E* **76** 046120
- [30] Centola D, González-Avella J C, Eguíluz V M and San Miguel M 2007 *J. Conflict Resolution* **51** 905–29
- [31] Vazquez F, Eguíluz V M and San Miguel M 2008 *Phys. Rev. Lett.* **100** 108702
- [32] Carro A, Vazquez F, Toral R and San Miguel M 2014 *Phys. Rev. E* **89** 062802
- [33] Diakonova M, Eguíluz V M and San Miguel M 2014 *Phys. Rev. E* **89** 062818
- [34] Diakonova M, Eguíluz V M and San Miguel M 2015 *Phys. Rev. E* **92** 032803
- [35] Holley R A and Liggett T M 1975 *Ann. Prob.* **3** 643–63
- [36] Vazquez F and Eguíluz V M 2008 *New J. Phys.* **10** 063011
- [37] Lancichinetti A, Radicchi F, Ramasco J J and Fortunato S 2011 *PLoS One* **6** e18961
- [38] Rosvall M and Bergstrom C T 2008 *Proc. Natl Acad. Sci. USA* **105** 1118
- [39] Gregory S 2010 *New J. Phys.* **12** 103018
- [40] Blondel V D, Guillaume J-L, Lambiotte R and Lefebvre E 2008 *J. Stat. Mech.* P10008
- [41] Fortunato S and Barthelemy M 2007 *Proc. Natl Acad. Sci. USA* **104** 36–41
- [42] Lancichinetti A, Fortunato S and Kertész J 2009 *New J. Phys.* **11** 033015
- [43] McDaid A F, Greene D and Hurley N 2011 arXiv:1110.2515
- [44] Suchecki K, Eguíluz V M and San Miguel M 2005 *Phys. Rev. E* **72** 036132
- [45] Fred A L N and Jain A K 2003 *Proc. 2003 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (Madison, WI, 18–20 June 2003)* p 128