

Natural Emergence of a Core Structure in Networks via Clique Percolation

A. Melka, N. Slater, A. Mualem, Y. Louzoun

Abstract—Networks are often presented as containing a “core” and a “periphery.” The existence of a core suggests that some vertices are central and form the skeleton of the network, to which all other vertices are connected. An alternative view of graphs is through communities. Multiple measures have been proposed for dense communities in graphs, the most classical being k -cliques, k -cores, and k -plexes, all presenting groups of tightly connected vertices. We here show that the edge number thresholds for such communities to emerge and for their percolation into a single dense connectivity component are very close, in all networks studied. These percolating cliques produce a natural core and periphery structure. This result is generic and is tested in configuration models and in real-world networks. This is also true for k -cores and k -plexes. Thus, the emergence of this connectedness among communities leading to a core is not dependent on some specific mechanism but a direct result of the natural percolation of dense communities.

Keywords—Networks, cliques, percolation, core structure, phase transition.

I. INTRODUCTION

REAL-WORLD networks as well as simulated graphs are characterized by multiple structural aspects. The first discussion of such a structure was the emergence of a giant connected component, followed by the connectedness of the majority of vertices, studied in 1960 by Erdős and Rényi [1], [2]. Networks can also have more local structures such as communities that have gained attention in recent years [3]–[6]. Those communities can either be loose and spanning the entire graph [7]–[10], or multiple dense subgraphs. Dense communities can be represented by structures, including among others k -cliques, k -plexes, or k -cores. Cliques are the most dense subgraph structure, in which each vertex must be connected to all the other vertices in the clique [11], [4]. k -plexes [12] are groups of vertices in which each vertex must be connected to all but $n - k$ vertices of the k -plex. For example, in a 3-plex of size 7, each vertex in the 7-vertices group has to be connected to only 4 other vertices, whereas, a clique of size 7 must have every vertex connected to the 6 other vertices. k -cores are even less restricted subgraphs [13]. In a k -core, all vertices must have at least k neighbors within the core. Every clique of size k is contained in the $(k - 1)$ -core and is a 1-plex of size k .

As is the case for single vertex connectedness, those dense communities can also percolate with high probability. Several models have been proposed for k -clique percolation. Following Derenyi et al. [14], [15], two k -cliques were considered adjacent if they shared $k - 1$ vertices. Using such a strict definition, the critical probability for the generation

of a giant component of k -cliques in the Erdős-Rényi (ER) networks is $P_c(k) = [(k - 1)N]^{-\frac{1}{k-1}}$, while the threshold for percolation in general is $P_c \propto N^{-1}$, where N is the total number of vertices in the graph. Their work has been broadened to less restrictive forms of percolation. For example, two k -cliques are considered to be adjacent if they share l vertices with $1 \leq l \leq k - 1$ [16], [17].

A third structural aspect of a network is the division between a set of dense connected high degree core vertices and sparse low degree periphery-vertices connected to other vertices through the core [18], [19]. Formally, core and periphery and dense communities are distinct concepts. The presence of a core does not necessarily imply the emergence of dense communities. However, there is *a priori* no reason for those communities to aggregate into a core. Nevertheless, communities and cores are correlated. Recently, Csermely and Uzzi have proposed that there is no clear discrimination between “networks modules” and “networks cores” [20]. Similarly, Rombach et al. [21] proposed that while networks can have both core and periphery structure and community structure, it would be desirable to study both and consider communities as “tiles” that overlap to produce a network core.

We here show that in configuration models [22] (such as ER and scale-free networks), real-world networks, and shuffled real-world networks, a percolation of all vertices belonging to k -cliques naturally emerges with a high probability, when such cliques are frequent. This percolation is shown using theoretical arguments and simulations. Although there is a gap between the edge density threshold for the existence of cliques and their percolation into a single connectivity component, this gap is very narrow and gets even narrower in large clique sizes. Thus, in practically every realistic network, the cliques percolate into a connected core. These results hold when we relax our constraint to all cliques with size above k and not just k -cliques since all vertices belonging to cliques larger than k are also in k -cliques. Moreover, our results hold even in less restrictive definitions, such as k -cores and k -plexes. For the purpose of our study, we will consider two k -cliques to be “neighbors” if an edge from a vertex in the first clique to a vertex in the second clique exists as illustrated in Fig. 1.

II. RESULTS

Recall that a clique is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent. To estimate whether all k -cliques are connected, we first determine the frequency of k -cliques in a configuration model graph. We then determine the probability that a randomly chosen vertex would be in a k -clique, and the

Bruno Melka is with the Bar-Ilan University, Israel (e-mail: bruno.melka@gmail.com).

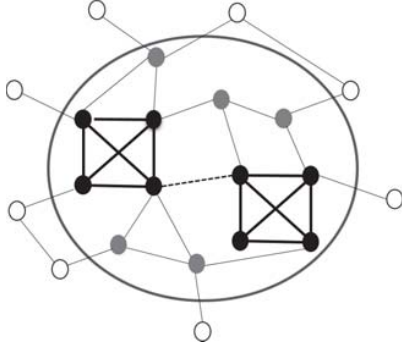


Fig. 1 Illustration of a 3-core (gray and black vertices inside the circle) that also contains two adjacent 4-cliques (black vertices connected through the dashed edge)

probability of two cliques of the same size to be neighbors (either connected by a common vertex, or having an edge connecting them). Finally, we determine the expected number of cliques of the same size to which a given clique is connected and compute the threshold for this number to be strictly greater than 1. Derivation of the results can be found in the Appendix. If the average number of other k -cliques neighboring each k -clique is strictly larger than 1, we can expect all cliques to percolate and all be connected to each other as the vertex number tends to ∞ .

We assume a configuration model graph with a general degree distribution, a number N of vertices, a probability $p(D)$ for a vertex to have a degree D , and further assume that the graph is fully connected. In this model, the degree distribution is given. Hence, the sum of all degrees $M = \sum_{i=1}^N D_i$ (also equal to twice the number of edges in the graph), has a narrow distribution around its expected value. With p_{ij} being the probability for two vertices i and j , with respective degrees D_i and D_j , to be connected, and $p_{ij|M}$ being the conditional probability, one obtains

$$p_{ij} \approx p_{ij|M} = \frac{D_i D_j}{M}. \quad (1)$$

The probability for k vertices with given degrees D_1, \dots, D_k greater than $k-1$ to be a clique is

$$\mathbb{P}(k \text{ given vertices to be a clique}) = \frac{\prod_{i=1}^k D_i^{k-1}}{M^{k(k-1)/2}}. \quad (2)$$

Summing over all possible degree values and approximating that there is no degree correlation leads to

$$\mathbb{P}(k \text{ vertices to be a clique}) = \frac{\left[\sum_{D \geq k-1} p(D) D^{k-1} \right]^k}{M^{k(k-1)/2}}. \quad (3)$$

To estimate the total number of cliques, one can choose $\binom{N}{k}$ combinations among all the vertices. We define $\mathbb{M}_k = \sum_{D \geq k} p(D) D^k$ as the modified k th moment of the degree distribution (modified since we only count from k). Therefore, the expected number \mathbb{N} of cliques of size k is

$$\mathbb{N} = \binom{N}{k} \frac{\mathbb{M}_{k-1}^k}{M^{k(k-1)/2}}. \quad (4)$$

To compute the average number of cliques neighboring a clique, we first compute the degree distribution of vertices in cliques, and then the probability of each edge leaving the clique to be connected to another clique. To compute if the neighbors of a vertex of degree D form a clique, we follow their degree distribution, which is distinct from the degree distribution of a randomly chosen vertex [7]. Assuming no degree correlation, this distribution is $\tilde{p}(D) = \frac{D p(D)}{\langle D \rangle}$ where $\langle D \rangle = \sum_D D p(D) = \frac{M}{N}$ is the average degree. For further use, we denote by \mathbb{Z} the probability for $k-1$ vertices to be a clique, following a given edge:

$$\mathbb{Z} = \left(\frac{N}{M} \right)^{k-1} \frac{[\mathbb{M}_{k-1} + p(k-2)(k-2)^{k-1}]^{k-1}}{M^{(k-1)(k-2)/2}}. \quad (5)$$

For a vertex of degree D not to be in a k -clique, none of the subsets of vertices of size $k-1$ within its neighbors should be a $(k-1)$ -clique. Since we use a configuration model, all edges are formed independently. We approximate that this is also the case for cliques, allowing us to multiply those probabilities:

$$\mathbb{P}(\text{vertex in } k\text{-clique} | D) = 1 - [1 - \mathbb{Z}]^{\binom{D}{k-1}}. \quad (6)$$

The probability for a vertex in a k -clique to have a degree D is then computed using Bayes theorem. $k-1$ neighbors of each vertex in the clique are connected to the other vertices in the clique, and there are k vertices in the clique. Therefore, for \mathbb{E} being the expected total number of neighbors for the clique (i.e., edges pointing outside the clique) to first order (ignoring edges pointing to the same vertex outside the clique), we obtain

$$\mathbb{E} = k \left[\sum_{D \geq k-1} \mathbb{P}(D | \text{vertex in } k\text{-clique}) D - (k-1) \right]. \quad (7)$$

Given \mathbb{E} the expected number of neighbors to a clique and the probability \mathbb{P} for a vertex to be in a clique of size k , the expected number of neighboring cliques of size k for a given clique of the same size is approximately $\mathbb{E} \cdot \mathbb{P}$. On average, the condition for all cliques of size k to be connected can then be argued to be $\mathbb{E} \cdot \mathbb{P}$ strictly larger than 1. Intuitively, as is the case in ER networks, if all vertices have, on average, slightly more than one neighbor, they are all connected. We actually apply the same kind of reasoning to k -cliques: if their average number of neighboring cliques is strictly greater than 1, then they are all connected, and percolation emerges.

To show that the thresholds for clique appearance and their percolation are actually very close, we first study the ER, $G(N, p)$, random graph, where N is the number of vertices and p is the probability for two vertices to be connected. The sum of all degrees is $M \approx pN^2$, and the average degree is $\lambda = \frac{M}{N} \approx pN$. The degrees have a Poisson distribution with an expected value λ . Their moment generating function is given by $f(t) = e^{\lambda(e^t - 1)}$. The variance is small and the

degrees for all vertices are centered around λ . Moreover, the relevant cliques are of k values much smaller than λ ; therefore, summing for all degrees greater than k is approximatively equal to summing for all degrees. As such, using our definition of the k th moment, we can approximate $\mathbb{M}_k \approx \lambda^k$, producing an estimate for the number of cliques of size k :

$$\mathbb{N}_{ER} \approx \binom{N}{k} \left(\frac{\lambda}{N}\right)^{k(k-1)/2} = \binom{N}{k} p^{k(k-1)/2}. \quad (8)$$

We can intuitively deduce the right part of this result since each connection is independent from the others and a clique of size k will have $k(k-1)/2$ edges, each of them having a probability p .

Assuming $N \gg k$, $\lambda \gg k$ and $\binom{N}{k} \approx \frac{N^k}{k!}$, one can closely approximate

$$\mathbb{N}_{ER} > 1 \iff p > \left[\frac{k!}{N^k} \right]^{\frac{2}{k(k-1)}}, \quad (9)$$

$$\mathbb{E}_{ER} \cdot \mathbb{P}_{ER} > 1 \iff p > \left[\frac{k!}{k^2 N^k} \right]^{\frac{2}{k(k-1)+2}}. \quad (10)$$

A more detailed derivation of those results can be found in the Appendix. Equation (9) represents the probability at which cliques of size k emerge, and (10) represents the probability at which those cliques percolate. Those two conditions are actually very close to each other as shown in Fig. 2. In the extreme case of all degrees being equal to λ , a regular graph is obtained, with the same results when replacing p by λ/N . The main difference is obtained in the 2-cliques, which are simply edges. As such, they appear as soon as there are edges, whereas their percolation only emerges for $p = 1/N$ as shown by Erdős and Rényi. However, when one requires the original graph to be connected even these differences disappear. Indeed, the green (medium gray) surface in the upper plot from Fig. 2 represents $p = 1/N$ and, therefore, the threshold at which the graph is connected. Cliques of size 2 percolate even a little before this threshold. Hence, to observe the percolation of cliques, one should actually look at the maximum between the green (medium gray) surface and the blue (dark gray) surface. Apart from this case, for higher values of k , the mere presence of cliques almost surely implies percolation: the blue (dark gray) surface is always above the green (medium gray) one. In graphs with a wider tail degree distribution, the two curves are even closer, as will be further discussed.

For power-law distributions, the probability of a vertex to have a degree D is $p(D) = \frac{CD^{-\alpha}}{N}$, where N is the number of vertices, C a normalization constant and α the slope of the distribution. For the purpose of our study, we chose $1.5 \leq \alpha \leq 3$. For values of α smaller than 2, the distribution has to be cut off with a maximum degree. $\langle D \rangle$ is the average degree. The constant C and the maximum degree D_{\max} are chosen such that: $\sum_{D=1}^{D_{\max}} CD^{-\alpha} = N$, and $\sum_{D=1}^{D_{\max}} CD^{-\alpha} D = N \langle D \rangle = M$. The average degree is much smaller than the one used for the ER graphs. Nevertheless, the gap between

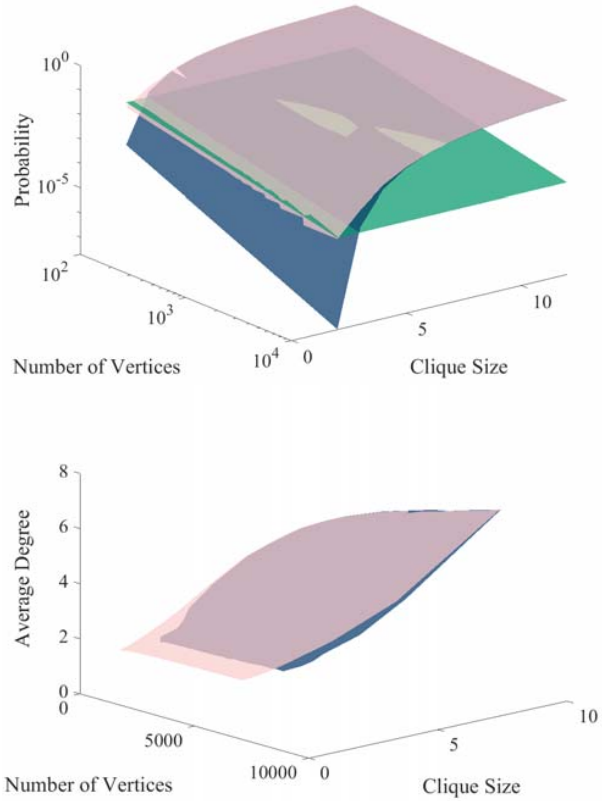


Fig. 2 Upper plot: Thresholds for clique appearance (blue/dark gray) and clique percolation (red/light gray) in terms of probability p for two vertices to be linked by an edge in an ER $G(N, p)$ random graph. Beyond a clique size of $k = 3$ the two surfaces are very close. The threshold for the emergence of cliques for $k = 2$ is close to the vertex percolation threshold. Thus, in connected networks, the values of p are typically on the green (medium gray) curve hence, much above the blue (dark gray) curve. The lower plot represents the same results for scale-free networks as a function of the average degree. The similarity between the two curves is even more pronounced in this case

the threshold for cliques to appear and their percolation is even narrower (see Fig. 2 for the theory and Fig. 3 for a comparison to simulations). Unlike the ER case, no closed form expressions could be derived and, therefore, we had to resort to numerical methods to evaluate the thresholds. The percolation in power-law degree distribution networks is the direct result of the presence of common hubs, as is the case for regular percolation, but even more stringent since the degree distribution within cliques is biased toward high degrees.

III. SIMULATIONS

To test the theoretical predictions and show that clique percolation is a universal feature, we tested for such a percolation on three types of networks: (A) configuration model simulated networks of different degree distributions, (B) real-world networks, and (C) shuffled real-world networks, where the degree distribution was maintained, and edges were randomly shuffled. For all studied graphs, we computed all cliques using the *Bron-Kerbosch* algorithm [23] and computed

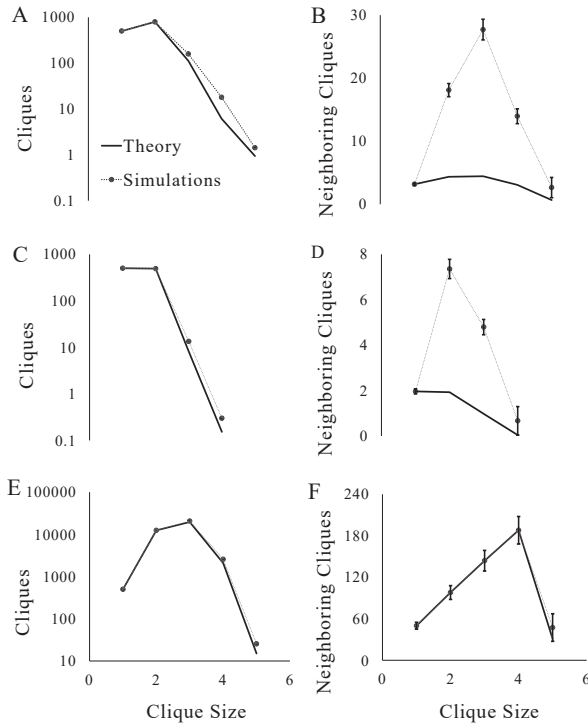


Fig. 3 Comparison between theory and simulations. Dots give the average over 500 simulations, and the full lines represent the theoretical results. The left column (A, C, E) shows the number of cliques of a given size. The standard error for this column is smaller than the marker size. The right column (B, D, F) shows the number of adjacent cliques to a random clique. While, for the clique number estimate, the theoretical result is a tight estimate, for the number of neighboring cliques, the theoretical result is a lower bound. The first row (A, B) is for power-law graphs with average degree 3 ($\alpha = 1.8$), and the second row (C, D), with average degree 2 ($\alpha = 2.2$). The third row (E, F) row is for ER graphs with an average degree of 50. All Graphs are for graphs with 500 vertices

the average number of cliques of each size in each real-world and simulated graph.

In the simulated networks, we sampled random cliques and counted how many neighboring cliques of the same size each clique had on average. In Erdős-Rényi graphs, the fit is tight. For power-law distributed graphs, there is a gap between theory and simulations, as expected by the higher overlap of cliques emerging around the high degree vertices in scale-free networks. Thus, clique percolation occurs for even lower average degree than expected by theory, further strengthening the claim that this is a universal feature. Note that we only studied percolation between cliques of the same size. Obviously, if we were to look at all cliques of size k and above, there would be even more neighbors and therefore percolation would happen for even lower average degree.

In all the results here, we assumed random pairing. However, multiple networks were reported to have degree assortativity. Such an assortativity could lead to a lower threshold for the emergence of cliques, and a gap between the threshold for the clique appearance and their percolation. We thus tested numerically whether the argument above still holds in assortative networks. To do so, we simulated ER

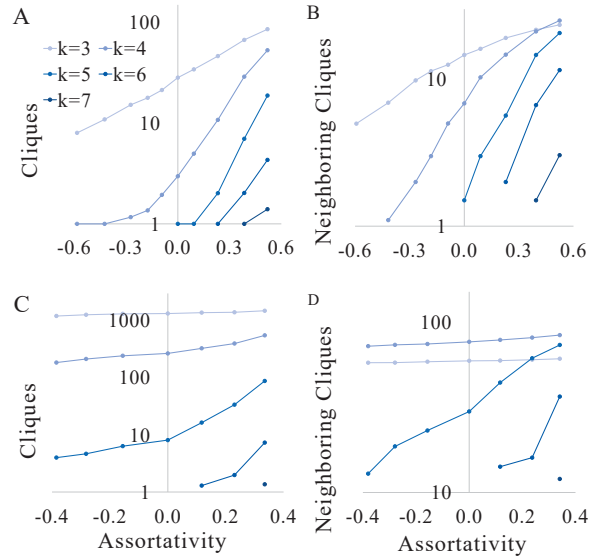


Fig. 4 Influence of assortativity (or disassortativity) on the number of cliques for different sizes. The first row (A, B) is for power-law graphs and the second row (C, D), for ER graphs. The left column (A, C) shows the number of cliques of a given size. The right column (B, D) shows the number of adjacent cliques to a random clique

and power-law networks and then, before computing the number of cliques and the average number of neighboring cliques, a certain fraction of the edges in those networks were reshuffled to achieve positive or negative assortativity. Edges were swapped only if this would increase (respectively, decrease) the assortativity. For $k = 1$ and $k = 2$, since the graph is connected and those cliques span over all the graph, the assortativity has no effect. For cliques of size above 2, positive or negative assortativity increases or decreases, respectively, the total number of cliques. This influence is more pronounced in the power-law distributed graphs (see Fig. 4). Still, whenever those cliques exist they percolate to form one connected component. We tested this feature by isolating the subgraph comprised only by nodes included in cliques of a given size and this subgraph was systematically connected. This compares to the theoretical threshold $\mathbb{E} \cdot \mathbb{P}$ (representing the expected number of neighboring cliques). Just like the total number of cliques grows with the assortativity, so does $\mathbb{E} \cdot \mathbb{P}$ and, as soon as cliques exist, $\mathbb{E} \cdot \mathbb{P}$ is greater than 1.

Real-world networks usually have a heavy tail. We checked the connectivity of the cliques of different sizes in some real-world networks. We calculated the probabilities for two cliques to be connected and compared the calculations to the actual fraction of vertices which are in the largest connected component of the subgraph consisting of only the k -cliques. As was the case for the ER network, in real-world networks, in the majority of studied network and clique sizes, when cliques are present, percolation occurs, and cliques form a single connected component [see Fig. 5 (A)]. To test that this is a random mechanism, and not related to any specific feature of the network, we shuffled the edges of the network, keeping the degree distribution, with an even clearer percolation [see

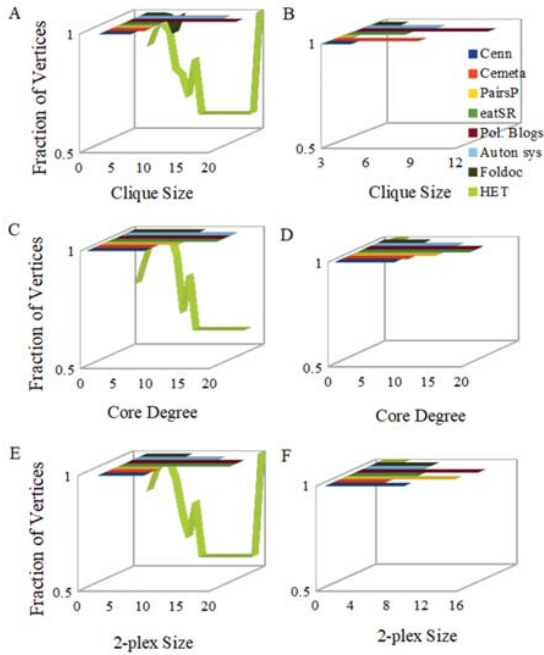


Fig. 5 Fraction of nodes in the main connectivity component as defined by k -cliques percolation (upper plots), k -cores (middle plots), and k -plexes (lower plots). The right columns are for the studied networks (A, C, E) and the left columns are for the scrambled version of the same networks (B, D, F). Each band is a different network. In the absence of cliques of a given size, no value is plotted. For all scrambled networks, and for the majority of real-world networks, all cliques, cores and plexes form a single fully connected component

Fig. 5 (B)].

A similar argument holds for other structures which are less stringent than cliques: k -plexes where $k = 2$ and k -cores. Note that the theoretical argument raised above for cliques works almost as is for k -cores and k -plexes. Thus, the natural emergence of a single connectivity component in such structures naturally is explained via similar reasons. For the majority of k values, all the vertices are part of one connected component, both for k -cores and for 2-plexes percolations [Figs. 5 (C) and (E)]. Again, when shuffling the networks, the result was maintained [Figs. 5 (D) and (F)]. The fact that the same results are obtained for the shuffled networks shows that the phenomena is related to the degree distributions and it is not a specific feature of the real-world networks.

IV. DISCUSSION

As mentioned above, multiple definitions have been proposed for the core and periphery structure. Some are based on cliques or other cohesive subgroups which are not cliques (for a review, see [24], [25]). We examined the vertices belonging to communities above a specific size k , and all the edges between them. We have shown that, when the graph itself is fully connected, these vertices percolate. We define the single connectivity component composed of these cliques as the network “core.” Unlike Derenyi et al. [14], we defined two cliques to be adjacent when they shared one vertex or when there is an edge between them.

Core and periphery structures have been argued to be the result of specific dynamical processes (such as assortativity [26]). At the static level, many real-world networks have been analysed in terms of core structure or the presence of communities, but it was broadly believed that those characteristics emerged due to the fact that some of the vertices had a dominant role and were therefore “strategically” positioned in essential locations inside the network. For instance, Wang et al. [27] state that essential proteins are placed as “hubs” in the network. Lin et al. [28] claim that those proteins would be part of cliques or cores. We demonstrated here that this does not have to be the case, and a core emerges naturally even in random networks. Strategic positioning may actually not be the consequence of the essentiality but the opposite. It is the position of a vertex that makes it essential and this process can happen randomly and is not based on any subjective characteristic. Kitsak et al. [29] also view cores or shells as more adequate to the spreading of information and “hubs” are key players but they assumed that those cores appeared according to the way people interact. We showed that, unless a process explicitly prevents the percolation of cliques, those communities will naturally form and percolate into a core. Instead of being a specific characteristic, core and periphery through clique percolation is actually the default setting. Note that historically, most of the studies on clique percolation were done on ER graphs with clique size rarely exceeding 4, explaining the observed gap between the emergence of cliques and their percolation as we mentioned above. In more realistic networks, this gap would be almost nonexistent.

Cores were used to determine the robustness of the network. Removing a vertex from the periphery has been argued not to affect the network, whereas removing a vertex from the core might jeopardize the whole structure. This is assuming that the core was formed through percolation between vertices alone. If, as we showed, the core emerges from the percolation of communities, then removing one vertex from the core will not affect the network since the rest of its community will ensure the connectivity. This definition of a core is much more robust than a definition based only on vertex degrees. This argument holds for both random and intended attacks. Similarly, Zhou et al. [30] showed that an intended process such as assortativity would actually decrease the robustness of the network and Xiao et al. [31] showed that re-wiring would actually improve the robustness. We suggest that for a high robustness, it indeed suffices to randomly build the network with the same degree distribution and naturally let cliques emerge and percolate into a core.

APPENDICES

A. Results Derivation

1) General Case:

Number of Cliques of Size k in a Graph: We assume a configuration model graph with a general degree distribution, a number N of vertices, a probability $p(D)$ for a vertex to have a degree D , and further assume that the graph is fully connected. In this model, the degree distribution is given. Hence, the sum

of all degrees $M = \sum_{i=1}^N D_i$ (also equal to twice the number of edges in the graph), has a narrow distribution around its expected value. With p_{ij} being the probability for two vertices i and j , with respective degrees D_i and D_j , to be connected, and $p_{ij|M}$ being the conditional probability, one obtains

$$p_{ij} \approx p_{ij|M} = \frac{D_i D_j}{M}. \quad (11)$$

The probability for k vertices with given degrees D_1, \dots, D_k greater than $k-1$ to be a clique is

$$\begin{aligned} \mathbb{P}(k \text{ given vertices to be a clique}) &= \prod_{i=1}^{k-1} \prod_{j=i+1}^k p_{ij} \\ &= \prod_{i=1}^{k-1} \left[\prod_{j=i+1}^k \frac{D_i D_j}{M} \right] \\ &= \frac{\prod_{i=1}^k D_i^{k-1}}{M^{k(k-1)/2}}. \end{aligned} \quad (12)$$

Summing over all possible degree value and approximating that there is no degree correlation leads to

$$\begin{aligned} \mathbb{P}(k \text{ vertices to be a clique}) &= \frac{\prod_{i=1}^k \left[\sum_{D \geq k-1} p(D) D^{k-1} \right]}{M^{k(k-1)/2}} \\ &= \frac{\left[\sum_{D \geq k-1} p(D) D^{k-1} \right]^k}{M^{k(k-1)/2}}. \end{aligned} \quad (13)$$

We define $\mathbb{M}_k = \sum_{D \geq k} p(D) D^k$ as the modified k th moment of the degree distribution (modified since we only count from k). Since one can choose $\binom{N}{k}$ combinations among all the vertices, the expected number \mathbb{N} of cliques of size k is

$$\mathbb{N} = \binom{N}{k} \frac{\left[\sum_{D \geq k-1} p(D) D^{k-1} \right]^k}{M^{k(k-1)/2}} = \binom{N}{k} \frac{\mathbb{M}_{k-1}^k}{M^{k(k-1)/2}}. \quad (14)$$

Note that for $k=1$, $\mathbb{N} = N$, the number of vertices in the graph and for $k=2$, $\mathbb{N} = M/2$, the number of edges in the graph.

Probability for a Vertex to Be in a Clique of Size k : To compute if a vertex is in a k -clique, we must first determine if the neighbors of a given vertex of degree D form a clique. We follow the degree distribution, which is distinct from the degree distribution of a randomly chosen vertex. Assuming no degree correlation, this distribution is $\tilde{p}(D) = \frac{Dp(D)}{\langle D \rangle}$ where $\langle D \rangle = \sum_D Dp(D) = \frac{M}{N}$ is the average degree. For further use, we denote by \mathbb{Z} the probability for $k-1$ vertices to be a clique, following a given edge:

TABLE I
LIST OF REAL-WORLD NETWORKS USED

Networks	Number of nodes	Average degree
Cenn	297	4.5
Foldoc	13356	13.7
PairsP	10617	12
eatSR	23218	26.3
Cemeta	453	9
Political blogs	1224	27.3
Autonomous systems	22963	4.2
High energy theory	7610	4.1

$$\begin{aligned} \mathbb{Z} &= \frac{1}{M^{(k-1)(k-2)/2}} \left[\sum_{D \geq k-2} \tilde{p}(D) D^{k-2} \right]^{k-1} \\ &= \frac{1}{M^{(k-1)(k-2)/2}} \left[\frac{\sum_{D \geq k-2} Np(D) D^{k-1}}{M} \right]^{k-1}, \quad (15) \\ &= \frac{\left(\frac{N}{M} \right)^{k-1}}{M^{(k-1)(k-2)/2}} \left[\sum_{D \geq k-2} p(D) D^{k-1} \right]^{k-1}, \\ &= \left(\frac{N}{M} \right)^{k-1} \frac{[\mathbb{M}_{k-1} + p(k-2)(k-2)^{k-1}]^{k-1}}{M^{(k-1)(k-2)/2}}. \end{aligned}$$

The probability for a vertex in a k -clique to have a degree D is then computed using Bayes theorem:

$$\mathbb{P}(D|\text{vertex in } k\text{-clique}) = \frac{\mathbb{P}(\text{vertex in } k\text{-clique}|D)p(D)}{\mathbb{P}(\text{vertex in } k\text{-clique})}, \quad (16)$$

where $\mathbb{P} = \mathbb{P}(\text{vertex in } k\text{-clique})$ is the probability for a vertex to be in a k -clique. Using the Law of Total Probability, we get

$$\mathbb{P} = \sum_{D' \geq k} \mathbb{P}(\text{vertex in } k\text{-clique}|D')p(D'). \quad (17)$$

The probability for a vertex of degree D to be in a k -clique is one minus the probability of this vertex not being in a k -clique. For this vertex not to be in a k -clique, none of the subsets of vertices of size $k-1$ within its neighbors should be a $(k-1)$ -clique. Each subset has a probability \mathbb{Z} to be a clique and, therefore, a probability $1 - \mathbb{Z}$ not to be a clique. Since we use a configuration model, all edges are formed independantly. We approximate that this is also the case for cliques, allowing us to multiply those probabilities:

$$\mathbb{P}(\text{vertex in } k\text{-clique}|D) = 1 - [1 - \mathbb{Z}]^{\binom{D}{k-1}}. \quad (18)$$

2) *ER Case:* We study the ER, $G(N, p)$, random graph, where N is the number of vertices and p is the probability for two vertices to be connected. The sum of all degrees is $M \approx pN^2$, and the average degree is $\lambda = \frac{M}{N} \approx pN$. The degrees have a Poisson distribution with an expected value λ . The variance is small and the degrees for all vertices are centered around λ . Moreover, the relevant cliques are of k values much smaller than λ ; therefore, summing for all degrees greater than k is approximatively equal to summing for all degrees. As such, using our definition of the k th moment,

we can approximate $\mathbb{M}_k \approx \lambda^k$, producing, with a first order expansion, an estimate for the number of cliques of size k :

$$\mathbb{N}_{\text{ER}} \approx \binom{N}{k} \left(\frac{\lambda}{N} \right)^{k(k-1)/2} \approx \frac{N^k}{k!} p^{k(k-1)/2}, \quad (19)$$

the probability for a vertex to be in a k -clique:

$$\begin{aligned} \mathbb{P}_{\text{ER}} &\approx 1 - \left[1 - \left(\frac{\lambda}{N} \right)^{(k-1)(k-2)/2} \right]^{\binom{\lambda}{k-1}}, \\ &\approx \binom{\lambda}{k-1} \left(\frac{\lambda}{N} \right)^{(k-1)(k-2)/2}, \\ &\approx \binom{\lambda}{k-1} p^{(k-1)(k-2)/2}, \end{aligned} \quad (20)$$

and the expected number of neighboring cliques:

$$\mathbb{E}_{\text{ER}} \approx k[\lambda - (k-1)]. \quad (21)$$

Combining those results leads us to

$$\begin{aligned} \mathbb{E}_{\text{ER}} \cdot \mathbb{P}_{\text{ER}} &\approx k[\lambda - (k-1)] \binom{\lambda}{k-1} \left(\frac{\lambda}{N} \right)^{(k-1)(k-2)/2}, \\ &\approx k^2 \binom{\lambda}{k} \left(\frac{\lambda}{N} \right)^{(k-1)(k-2)/2}, \\ &\approx k^2 \frac{\lambda}{N} \frac{N^k}{k!} \left(\frac{\lambda}{N} \right)^{k(k-1)/2}, \\ &\approx k^2 p \frac{N^k}{k!} p^{k(k-1)/2} = k^2 p \mathbb{N}_{\text{ER}}. \end{aligned} \quad (22)$$

From those calculations, we derive the thresholds for cliques of size k to emerge and for those cliques to percolate [(9) and (10)]. As a simple illustration, for cliques of size $k = 1$, which merely represents the vertices of the graph, we logically find $\mathbb{N} = N$ and their percolation appears for $p > 1/N$. For cliques of size $k = 2$, which represents the number of edges in the graph, $\mathbb{N} = pN^2/2 = M/2$. They appear for $p > 2/N^2$ and percolate for $p > 1/N\sqrt{2}$. Therefore, those cliques appear well before they percolate as shown in Fig. 2.

B. Real-World Networks and Methods

We studied eight real-world networks ranging from 297 to 23,218 vertices and an average undirected degree from 4 to 27. We used networks from various resources: biological networks, neural networks, and networks extracted from archives of citations, books and blogs [32] (see Table I for a description of the networks. The methods used were as follow:

k -cliques: We enumerated all cliques of different sizes on all networks. For each k , we generated a graph only from the vertices which belong to cliques of sizes k and above. We computed the fraction of vertices (of the subgraph consisting of the cliques of sizes k and above only) belonging to the largest connected component.

k -plexes and k -cores: We enumerated all the 2-plexes of all sizes using a C++ implementation of the Wu and Pei

algorithm [33]. We then measured the fraction of vertices in the largest connected component of the graph composed of only the 2-plexes of sizes equal or larger than k . For the k -cores, we used C++ implementation based on the Boost library which lists every vertex in the graph and is assigned several of its k -cores. We checked the fraction of vertices in the largest connected component of each k -core graph.

Networks shuffling: For each real-world network, we have generated a random network with the same number of vertices and the same degree distribution as the original network. This algorithm is sometimes called “local rewiring” [34]. The shuffled networks were obtained by iteratively choosing a pair of random edges and switching their destination. For example, the edges (1,2) and (3,4) in the original network were switched to edges (1,4) and (2,3) in the shuffled network. We checked that the newly obtained edges are not self-edges or overlapping edges. If the new candidate edges were found to be self-edges or overlapping edges, we canceled the switch and chose a different pair of edges to switch. There were a few networks in which double or self-edges could not be avoided, mainly in cases where vertices had a very large degree. This had a very minor effect on the degree or clique distribution.

REFERENCES

- [1] P. Erdős and Al. Rényi, Publ. Math. 6, 290 (1959).
- [2] P. Erdős and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. 5, 17 (1960).
- [3] M. Porter, J. Onnela, and P. Mucha, Not. Am. Math. Soc. 56, 1082 (2009).
- [4] R. Mokken, Qual. Quant. 13, 161 (1979).
- [5] J. Duch and A. Arenas, Phys. Rev. E. 72, 027104 (2005).
- [6] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, Proc. Nat. Acad. Sci. USA. 101, 2658 (2004).
- [7] M. E. J. Newman, SIAM Rev. 45, 167 (2003).
- [8] M. E. J. Newman and Michelle Girvan, Phys. Rev. E. 69, 026113 (2004).
- [9] A. Clauset, M. E. J. Newman, and C. Moore, Phys. Rev. E. 70, 066111 (2004).
- [10] M. E. J. Newman, Proc. Nat. Acad. Sci. 103, 8577 (2006).
- [11] R. Luce, Psychometrika. 15, 169 (1950).
- [12] S. Seidman and B. Foster, J. Math. Soc. 6, 139 (1978).
- [13] S. Dorogovtsev, A. Goltsev, and J. Mendes, Phys. Rev. Lett. 96, 040601 (2006).
- [14] I. Derényi, G. Palla, and T. Vicsek, Phys. Rev. Lett. 94, 160202 (2005).
- [15] G. Palla, I. Derényi, and T. Vicsek, J. Stat. Phys. 128, 219 (2007).
- [16] J. Fan and X. Chen, Europhys. Lett. 107, 28005 (2014).
- [17] M. Li, Y. Deng, and B. Wang, Phys. Rev. E. 92, 042116 (2015).
- [18] S. Borgatti and M. Everett, Soc. Networks, 21, 375 (2000).
- [19] P. Holme, Phys. Rev. E. 72, 046111 (2005).
- [20] P. Csermely, A. London, L. Wu, and B. Uzzi, J. Comp. Net. 1, 93 (2013).
- [21] P. Rombach, M. Porter, J. Fowler, and P. Mucha, SIAM Rev. 59, 619 (2017).
- [22] M. Catanzaro, M. Boguná and R. Pastor-Satorras, Phys. Rev. E. 71, 027103 (2005).
- [23] C. Bron and J. Kerbosch, Commun. ACM, 16, 575 (1973).
- [24] J. Scott, Soc. Networks (2017).
- [25] S. Wasserman and K. Faust, Camb. Univ. Press, 8 (1994).
- [26] M. Piraveenan, M. Prokopenko, and A. Zomaya, Euro. Phys. J. B. 70, 275 (2009).
- [27] J. Wang, M. Li, H. Wang, and Y. Pan, IEEE T. Comp. Bio. 9,1070 (2012).
- [28] C. Lin, H. Juan, J. Hsiang, Y. Hwang, H. Mori, and H. Huang, J. Proteome Res. 8, 1925 (2009).
- [29] M. Kitsak, L. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. Stanley, and H. Makse, Nat. Phys. 6, 888 (2010).
- [30] D. Zhou, H. Stanley, G. D’Agostino, and A. Scala, Phys. Rev. E. 86, 066103 (2012).
- [31] S. Xiao, G. Xiao, T. Cheng, S. Ma, X. Fu, and H. Son, Europhys. Lett. 89, 38002 (2010).
- [32] L. Adamic, N. Glance, Proceedings of the 3rd International Workshop on Link Discovery (2005).

- [33] N. Du, B. Wu, L. Xu, B. Wang, X. Pei, Proceedings of the Data Mining Workshops (2006).
- [34] S. Maslov, K. Sneppen, A. Zaliznyak, Physica A, 333, 529 (2004).