Physica A 402 (2014) 330–343

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

The impact of network characteristics on the diffusion of innovations

Renana Peres *

School of Business Administration, Hebrew University of Jerusalem, Jerusalem, 91905, Israel

HIGHLIGHTS

- Directly studies the dependence of new product diffusion on network topology.
- Uses agent-based models on 160 networks for monopoly and duopoly markets.
- Generates networks using random graphs with a planted partition.
- Average-degree and high-degree hubs enhance diffusion.
- Clustering negatively affects diffusion.

ARTICLE INFO

Article history:
Received 8 August 2013
Received in revised form 23 January 2014
Available online 12 February 2014

Keywords:
Clustering
Average degree
Social hubs
Agent-based models
Diffusion of innovations
Word of mouth

ABSTRACT

This paper studies the influence of network topology on the speed and reach of new product diffusion. While previous research has focused on comparing network types, this paper explores explicitly the relationship between topology and measurements of diffusion effectiveness. We study simultaneously the effect of three network metrics: the average degree, the relative degree of social hubs (i.e., the ratio of the average degree of highly-connected individuals to the average degree of the entire population), and the clustering coefficient. A novel network-generation procedure based on random graphs with a planted partition is used to generate 160 networks with a wide range of values for these topological metrics. Using an agent-based model, we simulate diffusion on these networks and check the dependence of the net present value (NPV) of the number of adopters over time on the network metrics. We find that the average degree and the relative degree of social hubs have a positive influence on diffusion. This result emphasizes the importance of high network connectivity and strong hubs. The clustering coefficient has a negative impact on diffusion, a finding that contributes to the ongoing controversy on the benefits and disadvantages of transitivity. These results hold for both monopolistic and duopolistic markets, and were also tested on a sample of 12 real networks.

1. Introduction

The social influence processes that take place in a given network are shaped and affected by the network's topological characteristics. In this paper, we study how the topological or structural characteristics of a social network influence new-product diffusion in that network, in terms of speed of diffusion and the number of adopters.

Classical works in diffusion, focusing on the flow of information among individuals, assumed a fully connected market [1]. More recently, the literature has begun to acknowledge the role of network topology in social influence processes, exploring

* Tel.: +972 2 5883073.
E-mail address: peresren@huji.ac.il.

http://dx.doi.org/10.1016/j.physa.2014.02.003
0378-4371/

diffusion in topologies such as small world networks [2] and scale-free networks [3]. Empirical studies have explored how diffusion is influenced by aspects of network structure, including weak, long-distance ties [4] and the existence of social hubs [5], and have examined how network structure affects the performance of marketing strategies such as new-product seeding [6,7].

Despite this interest, to our knowledge, the direct impact of topological network metrics on diffusion has not been studied. Specifically, there has not been a systematic assessment, carried out across multiple networks, testing the *simultaneous* and *direct* impact of multiple topological metrics on the magnitude and speed of diffusion. Although some comparative studies have been conducted, most of them compare network types (e.g. Ref. [8]), referring to the comprehensive set of properties characterizing each network rather than isolating the specific role of each structural dimension.

In this paper, we conduct a methodological investigation of the impact of network topology on the diffusion of a new product to the market. Specifically, we focus on the following structural metrics: the average degree, the relative degree of social hubs (i.e., the ratio between the average degree of the most-connected nodes and the overall average degree), and the clustering coefficient. We apply a graph-theory procedure, a variation on the l-partition model, which uses random graphs with a planted partition, which has so far not been used in diffusion research, and use it to generate 160 networks, with a large range of values of the investigated metrics. We conduct agent-based simulations of new product diffusion in these networks, both in a monopoly and under duopolistic competition. We test the relative influence of each structural metric on the effectiveness of diffusion, measured as the Net Present Value (NPV) of the number of adopters, and controlling for the diffusion parameters and the networks' degree of separation.

Our main findings are:

1. Among the investigated metrics, the average degree and the relative degree of hubs have a strong positive impact on diffusion. The latter result is interesting in light of the controversy on the contribution of social hubs [9,10].
2. The clustering coefficient has a negative impact on diffusion. This result is in line with previous works comparing network types; however, it isolates the role of clustering from that of other topological network metrics. In addition, this finding contributes to an ongoing discussion on the benefits and disadvantages of transitivity (that is, the likelihood that the other nodes connected with a node are also connected to one another), of which clustering is a measure [11], implying the possible drawbacks of transitivity in the context of diffusion.

This paper offers three main contributions: First, it measures the *direct impact* of structural parameters on diffusion. We vary three major network metrics, and run a multivariate regression to explore simultaneously their *relative* roles. Second, we use a network generation method that has not been used so far in diffusion research, to create a set of networks with a wide range of values for the various metrics we examine, without the need to use networks of different types. This is a variation on the l-partition model from graph theory, which is novel in two aspects: it can generate networks which are more realistic than the classical l-partition model, and, more important, it enables to create networks with *pre-specified metrics*, including both degree distribution and clustering coefficient. Third, the agent-based simulation enables us to represent real-life diffusion scenarios by (i) considering both a monopoly and a competitive market; (ii) using a cascade agent activation model [12]; this is the individual-level analog to the Bass diffusion model [1], which is the standard model used to describe diffusion processes; and (iii) evaluating the effectiveness of the diffusion process by measuring the NPV of the number of adopters, which reflects both the reach of the diffusion process and its speed. Previous studies used such simulations, but did not focus on isolating the roles of specific topological metrics in the diffusion process.

The rest of this paper is organized as follows: In Section 2 we review the literature and describe the topological metrics we use and their anticipated influence. Section 3 describes the network generation procedure. Section 4 describes the agent-based model. The results and conclusions are presented in Sections 5 and 6, respectively.

2. Network structure and diffusion propagation

Social influence processes are strongly affected by the topology of the network in which they take place. Thus far, most studies focusing on the relationships between network topologies and social influence have been non-comparative in nature, with each paper focusing on a single network type. To the extent that comparison was done, it focused on comparing performance across different types of networks, and was mostly theoretic without measuring actual diffusion or information flow (see Refs. [13,14]; see Ref. [8] for an exception of an empirical study). For example, in small world networks, which are "rewired" lattice graphs in which several lattice ties are replaced with random connections, information flow and influence processes are expected to be more rapid than in regular lattice graphs, due to the "shortcuts" between nodes[1] [14,15]. Likewise, information flow in fully random graphs is expected to be rapid [16,17].

These studies provide important insights with regard to the influence of network structure on information flow. However, a specific network type usually binds several topological dimensions: For example, small-world networks combine both high clustering and short path length; most random graphs have a Poissonian degree distribution (with some exceptions;

[1] The network literature uses a variety of terms to describe network members and their connections. We use the term "network member" when talking about the real individuals in the network, "node"—to describe this person in the theoretical context, and "agent" when speaking about the agent-based model. In all contexts, we refer to the connections among network members, nodes, or agents as "ties".

see Ref. [16]). In scale-free networks the clustering depends on the network size [13]. Thus, it is hard to isolate the relative role of each structural dimension on the basis of an overall comparison of different types of graphs. For example, if diffusion in a lattice graph is slower than in a random graph, is this because of the lower clustering of the random graph, or perhaps the variability in the degree of the nodes? Do the hubs in scale-free networks generate faster diffusion in comparison with other graphs, although the average degree in scale-free networks is very low? These questions are still unanswered.

The goal of this paper is to study systematically the influence of structural factors on the speed of diffusion. We focus on three structural dimensions: average degree, relative degree of social hubs, and clustering coefficient. While the network literature suggests numerous network metrics, we focus on these three metrics since (i) they are network level metrics that measure global properties rather than local characteristics, (ii) they are widely used to characterize networks in the literature [13,18,19], and, (iii) as we further show, they can be specified independent of each other using our network generation algorithm.

2.1. Average degree

The degree of a node is the number of ties it has with other nodes. The average degree of a network is the average of the degrees of all the nodes in the network, and can be considered as a metric of the network's level of connectivity. Empirical comparisons of the average degrees of real-life networks indicate values ranging from $\sim$7 in some neural networks to $<$113 in a networks of film actors [19,20]. For theoretical networks, the average degree can often be derived from the networks' creation procedure: For example, for a regular lattice or for a Watts–Strogatz network, where all nodes have an equal number of k ties, the average degree is k, and a random Erdős–Renyi graph of N nodes with a tie probability p will have an average degree of $N \cdot p$.

2.2. Relative degree of social hubs

Social hubs are nodes with a high degree in relation to the degrees of other nodes. Not all networks have social hubs. A regular lattice does not have any. In random graphs and small world networks, the degree distribution is Poissonian [13], and social hubs do not differ much from the other members of the network. Social hubs are most prevalent in scale-free networks, where degree distribution follows a power law, with most of the nodes having a small number of ties, and a small number of nodes having an extremely high degree [13,21].

The contribution of social hubs to diffusion processes is a subject of ongoing debate. Some studies argue that hubs enhance the spread of information [22] and the speed of diffusion [6], and hence it is beneficial to target these individuals when attempting to introduce new products or ideas into a network [7]. Other studies claim that the role of hubs is complicated and depends on the level of contagion in the system [23], as well as on these individuals' inherent propensity to adopt [5]. Watts and Dodds [9] suggest that in many cases, the large mass of less-connected units in a network determines the speed and magnitude of social influence [9].

Here, we contribute to this discussion by systematically varying the level of the relative degree of hubs across the networks we generate, and testing the influence of this variable, while controlling for other topological metrics. We measure the relative degree of social hubs as the *degree ratio* of the average degree of the top 10% most connected nodes to the overall average degree.

2.3. Clustering coefficient

The clustering coefficient serves as a measure of a network's transitivity, that is, the likelihood that if nodes a and b are connected to each other, and nodes b and c are connected to each other, then nodes a and c are also connected. In other words, the clustering coefficient indicates the likelihood that a person in a given network is a friend with the friends of his or her friends. Clustering can be measured either locally for each node – counting the number of its ties, and seeing how many of them are connected to each other – or globally, counting the numbers of "triangles" relative to the number of "open triples", where an open triple is a single node with ties to two other nodes. For example, if nodes a, b and c are all connected to each other, they form a single triangle, and three open triples: abc, bac, acb. Herein we use Newman's [19] definition of a global clustering coefficient [19]:

$$C = \frac{3 \times \text{number of triangles in the network}}{\text{number of open triples}}. \tag{1}$$

Clustering is strongly dependent on network type, and can also vary to some extent across networks of the same type. For a one-dimensional lattice arranged in form of a ring, where each node is connected to its k nearest neighbors, the clustering coefficient is $C = \frac{3(k-2)}{4(k-1)}$, which approaches $C \sim 3/4$ for large values of k. In a random graph, clustering tends to be much lower, given by the approximation of k/N, where N is the network size, and k is the average degree of the network. The clustering coefficient of a small-world Watts–Strogatz network is between these values [14] and depends on the value of p, the rewiring probability of the regular lattice $C \sim \frac{3(k-1)}{2(2k-1)}(1-p)^3$ [24]. In scale-free networks generated by the Albert–Barabási

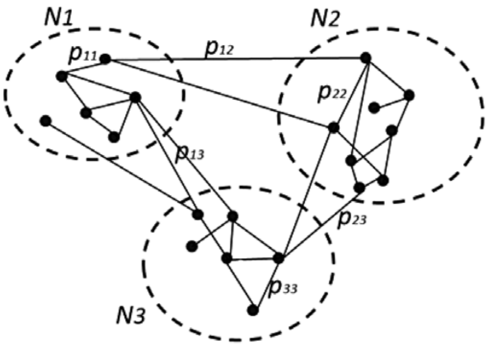

Fig. 1. A random graph with a planted partition, 3 bins, unequal probabilities.

procedure, clustering is higher than that in a random graph, but lower than that in a small world graph, and is given approximately by $C \sim N^{-0.75}$ [13]. The strong dependence of clustering on network type can be an obstacle for systematically testing how clustering affects diffusion, since it is hard to determine whether differences result from the clustering or from other characteristics that are typical of specific network types, such as level of randomness, path length, etc.

Network theory is split with regard to whether transitivity is beneficial for network processes. In some contexts, a node, that is, an individual in the network, is better off not closing a triad, but rather forming ties with a new node; in other contexts, such as when redundancy and multiple points of influence are needed, individuals gain from closing a triad and increasing the network's transitivity (see Ref. [11] for review). Here, we shed some light as to the role of transitivity in the context of diffusion, by systematically varying the clustering coefficient, and testing directly its effect on diffusion, controlling for other topological metrics.

As mentioned above, the network literature suggests various additional network level metrics that can potentially influence diffusion processes (e.g. average path length, diameter, density), as well as many local, node-level metrics such as the local centrality measures (see Ref. [25] for a list of metrics). We focus here on network-level properties, choosing the three metrics that are most commonly used in network literature [13,18,19]. These metrics represent independent topological network dimensions: the level of connectivity (represented by the average degree), degree homogeneity (represented by the relative degree of hubs), and transitivity, measured by the global clustering coefficient. Also, the network generation procedure suggested below enables to vary these metrics independent of each other and to generate networks with pre-selected values of these metrics. An additional dimension of importance, which is not represented by these metrics relates to the average distance between nodes. Unlike the three focal metrics we do not pre-specify or vary it systematically, but rather measure it (using the average path length metric) and use it as a control variable in our analysis.

3. Generating the networks

To test how network metrics influence diffusion, it is necessary to generate a large number of networks, with a wide range of values for each metric under investigation. To enable an unbiased estimation of the impact of each metric, a proper experimental design which will vary the metrics systematically and independently is needed. Since the standard network types (random, lattice, small-world, etc.) largely dictate dependencies among the network metrics, we introduce here a network generation method, based on merging several random graphs, which can generate networks with pre-selected values for the average degree, relative degree of social hubs, and clustering coefficient. This method is a variation on the "random graphs with a planted partition" suggested by graph theorists for dealing with problems such as graph partitioning and community detection [26,27].[2] Although it was discussed in various contexts, to the best of our knowledge this procedure has not yet been applied for generating social networks with specified metrics and has not been used in diffusion simulations.

Assume a market of size N. For simplicity, ties are symmetric, i.e., if node a is connected to b, then b is also connected to a. The N nodes are organized into three separate bins (denoted bins 1, 2 and 3) of sizes N_1, N_2, and N_3, respectively. The probability that a customer from bin i and a customer from bin j are connected is p_{ij}. For three bins, there are six probabilities: $p_{11}, p_{22}, p_{33}, p_{12}, p_{13}, p_{23}$, as illustrated in Fig. 1. If all these probabilities are the same, the graph is a standard random graph. Manipulating the probabilities allows one to create random networks with desired values for the average degree, degree ratio and clustering as follows:

The average degree of the nodes in bin i is given by $D_i = p_{ii} \cdot (N_i - 1) + p_{ij} \cdot N_j + p_{ik} \cdot N_k$, where $i, j, k = 1, 2, 3$ and $i \neq j \neq k$. The term $p_{ii} \cdot (N_i - 1)$ is the average number of ties that the node has with other nodes in bin i (assuming a node cannot connect to itself), and the two other terms are the number of ties that the node has with nodes in bins j and k,

[2] We thank Michael Krivelevich for several useful discussions on this topic.

respectively. Hence, the overall average degree is $D = \sum_i (D_i \cdot N_i/N)$. If we arbitrarily define bin 1 as the bin of hubs and set p_{11} to be higher than the other probabilities, then we can define the relative degree of hubs as D_1/D.

To calculate the clustering coefficient, we need to calculate the number of triangles and the number of open triples. The number of triangles is given by the following expression:

$$\text{Triangles} = \sum_{i=1}^{3} \binom{N_i}{3} p_{ii}^3 + \sum_{i=1;\, j,k\neq i}^{3} \binom{N_i}{2} p_{ii}(p_{ii}^2 N_j + p_{ik}^2 N_k) + N_1 N_2 N_3 p_{12} p_{13} p_{23}. \tag{2}$$

The first sum is the number of triangles in which all nodes belong to the same bin. The second sum is the number of triangles in which two nodes are in the same bin and one is in a different bin: If $i = 1$, for example, we first choose a pair of nodes in bin 1 (there are $\binom{N_1}{2}$ such pairs); their probability to be connected is p_{11}. Taking randomly a node in bin 2 (there are N_2 such nodes), the probability that both nodes of bin 1's pair will be connected to it is p_{12}^2; the same logic applies for bin 3. The third term is the number of triangles in which each node is in a different bin.

The number of open triples is calculated in a similar way. A unit in bin i can be in an open triple with either two other units in i, two units in j, or two units in k, or with one in i and one in j, one in i and one in k, or one in j and one in k. Since there are N_i units in bin i, the number of open triples that contain a node from bin i is given by:

$$\text{OpTr}_i = N_i \left(\binom{N_i - 1}{2} p_{ii}^2 + \binom{N_j}{2} p_{ij}^2 \binom{N_k}{2} p_{ik}^2 + (N_i - 1) N_j p_{ii} p_{ij} + (N_i - 1) N_k p_{ii} p_{ik} + N_j N_k p_{ij} p_{ik} \right). \tag{3}$$

The clustering coefficient is then calculated as $C = \frac{3 \times \text{Triangles}}{\text{OpTr}_1 + \text{OpTr}_2 + \text{OpTr}_3}$.

Given values of the average degree, relative degree of hubs, and clustering coefficient, we can find a set of probabilities that satisfy these values. Finding the probabilities is done through a numerical optimization software (here, the Excel Solver was used), which looks for a set of 6 probabilities that provide the desired metric values. Note that this procedure does not guarantee a solution for every degree/ratio/clustering combination, or that an obtained solution is unique (since we have 6 probabilities and 3 equations). However, since the goal is to generate networks with a desired structure and not to obtain a specific network, this procedure is adequate for our purposes. Once the probabilities are determined, a random number generator is used to create the networks corresponding to these probabilities.

This procedure was used to generate 160 networks, using $N = 1000$, $N_1 = 100$, $N_2 = 450$, and $N_3 = 450$. The networks span a wide range of values in each of the three network metrics. The range of values for each metric was chosen to cover the range observed in real social networks, as described in experimental social network papers (see Refs. [13,18,19] for several comparative tables). Although these papers analyze a small number of networks which do not aim to represent the entire social network universe, we used the range they provide to guide our choice of ranges for the metrics in the generated networks. Thus, the average degree in our networks ranges from 2 to 50 (average is 24.5), the clustering coefficient ranges from 0.01 to 0.48 with an average of 0.21, and the relative degree of hubs (i.e., the ratio between the average degree of the 10% most connected nodes and the overall average degree) ranges from 1.09 to 6.7 with an average of 3.96. Note, that we limited the upper bound of the average degree to 50, since our network size is limited, and a high average degree will limit the range of clustering coefficients and relative hubs degree we could use.

To generate each of the 160 networks, a combination of average degree, clustering coefficient and relative degree of hubs was chosen from within the above ranges. To choose the values, we used Latin hypercube sampling (LHS). In the LHS experimental design [28,29] the range of each variable is divided into non-overlapping intervals (160 in our case[3]) on the basis of equal probability. One value from each interval is selected at random with respect to the probability density in the interval (with no prior information, we assumed uniform density). Then, the 160 values thus obtained for each metric are coupled in a random manner (equally likely combinations) to create 160 combinations.

Latin hypercube sampling is used frequently in response surface methodology studies in chemical engineering and aeronautics [30]. Due to the efficient manner in which it stratifies across the range of each sampled variable, this is often the preferred sampling procedure in analyses where the independent variables are continuous and their distribution is unknown. These conditions apply in our case—there is not enough empirical evidence as to the distribution of each of the three network metrics across networks (large scale comparative statistics of network parameters is hardly available), and we wanted to vary the metrics independently and cover the entire range. LHS is favored over a standard full factorial design, since it allows a better sampling of the parameter space of each variable, and is favored over than standard Monte Carlo due to its stratified nature. It leads to low bias and high fit in regression analysis, which is the analysis used in this paper [31]. The distribution of parameters across the networks is described in Fig. 2. In addition, the average path length for each network was measured (e.g. the average distance over all pairs of nodes). The average path length ranges from 1.71 to 186, with an average of 5.33. It was used as a control variable in our analysis.

Fig. 3 illustrates the network structure of a network with an average degree of 7.5, relative degree of hubs of 6.67, and a clustering coefficient of 0.45. The probabilities are: $p_{11} = 0.49$, $p_{22} = 0.0001$, $p_{33} = 0.012$, $p_{12} = 0.013$, $p_{13} = 0.013$, $p_{23} = 0$.

[3] The number was chosen from computational considerations.

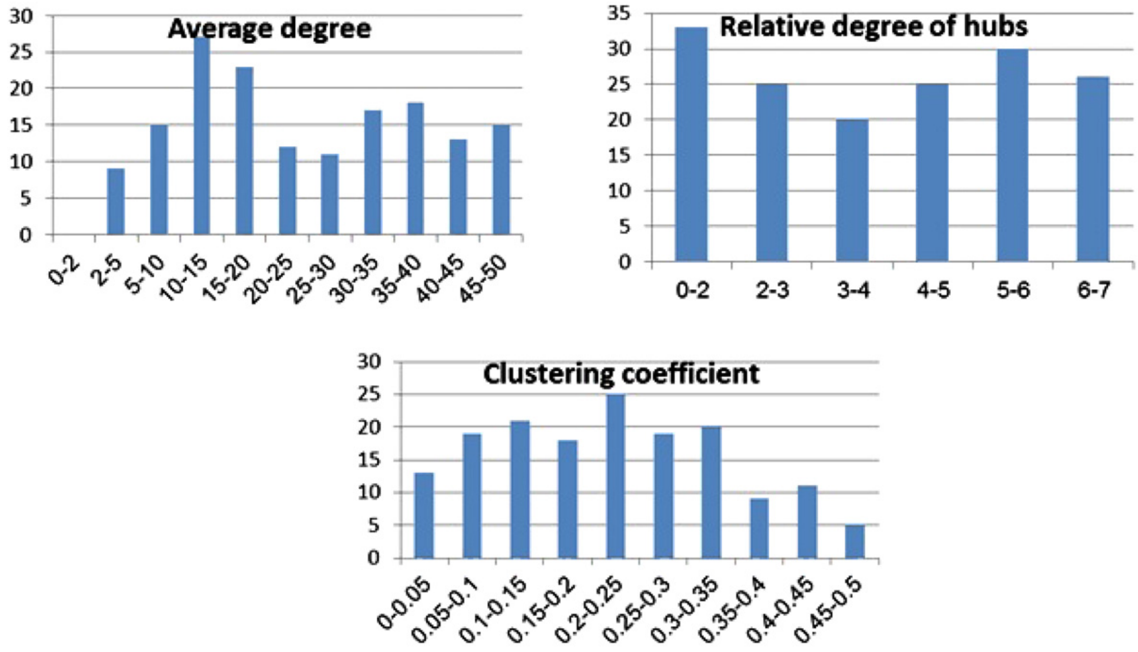

Fig. 2. The distribution of parameters across the 160 networks.

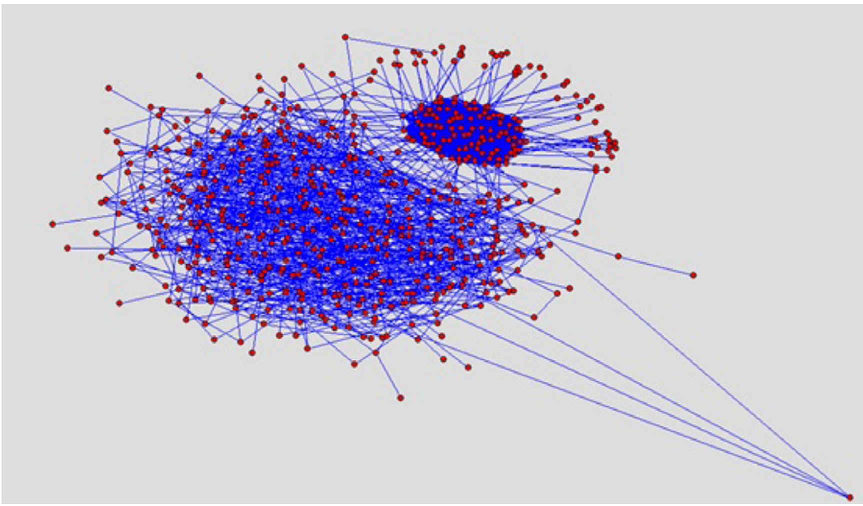

Fig. 3. An example of a network connectivity generated using the network generation procedure. Average degree = 7.5, relative degree of hubs = 6.67, clustering coefficient 0.45. $p_{11} = 0.49,\ p_{22} = 0.0001,\ p_{33} = 0.012,\ p_{12} = 0.013,\ p_{13} = 0.013,\ p_{23} = 0.$

The 160 networks we use here span a wide range of degree distributions, including bell shaped curves, as well as distributions resembling power law (see Fig. 4 for illustration). While one can think of alternative network generation algorithms that can create pre-defined degree distributions [32], or generate networks with a large number of parameters, from which one can post-hoc choose the networks with the desired parameter combinations [33], this method is unique since it can generate networks with a pre-selected *combination* of both *degree metrics* as well as *clustering coefficient*. Thus, it enables to systematically and independently vary these metrics in order to build methodological experimental designs.

A noteworthy question with respect to this procedure is to what extent it represents networks that resemble real social networks. This question was discussed in the context of community detection algorithms, where partitioned graphs are usually used as benchmarks [27]. Although the community structure is a characteristic of social networks, pure planted partition graphs (with equal bins, and only two probabilities—p_{in} for within bin connectivity, and p_{out} for out of bin connectivity) were criticized as being non-realistic in the sense they do not show degree heterogeneity, and have a low clustering

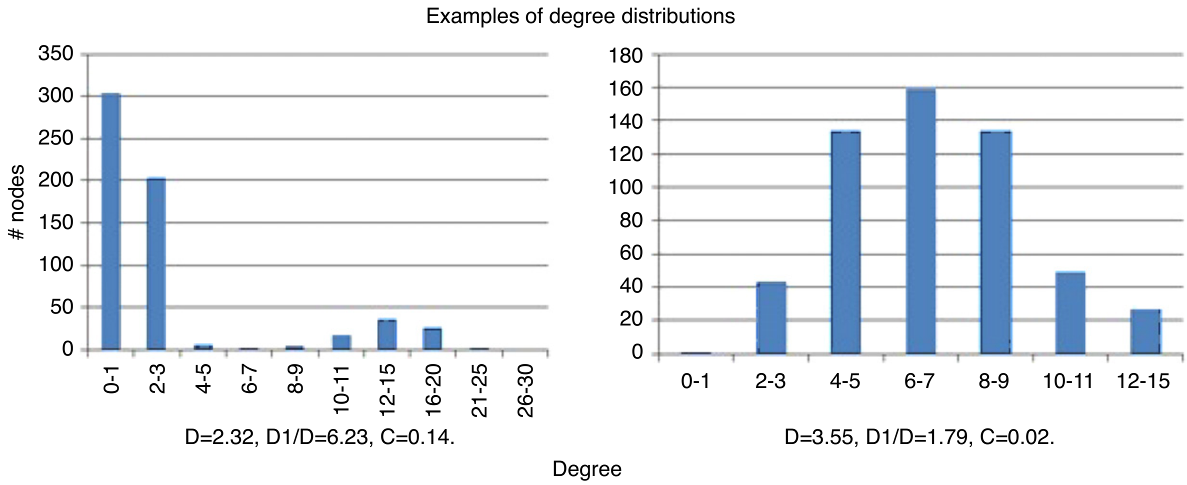

Fig. 4. Example of degree distribution of two of the 160 networks.

coefficient [24,34]. Various improvements were suggested to overcome these limitations [35]. The variation on the planted partition graph suggested here overcomes some of the limitations of the pure planted partition graphs, as follows:

- The bins vary in size, and the probabilities p_{ii} and p_{ij} vary across bins. This creates degree heterogeneity, as recommended by Ref. [35]. The degree-distributions remain in the realm of Poissonian mix, but, as indicated by Fig. 4, certain combinations of probabilities generate skewed distributions, similar to power law. Most importantly, the degree distributions generated by this procedure allow for social hubs, which is the social phenomenon we focus on.
- The global clustering coefficient is not small as in random graphs but rather covers a wide range of values. The low clustering has been considered a major drawback of random graphs and an open issue in benchmarking methods and network generation algorithms [33,35].
- The range of average degrees and clustering coefficients corresponds to the range found in empirical studies [13,18,19]. The average path length (measured as a control, but was not part of the experimental design) also corresponds to that of real networks.
- This method enables generating hierarchical structures such as communities within communities, through a manipulation of the bin sizes and connection probabilities. Although not used here, this is a desired feature of real networks [27,35].

Thus, our method aims to balance the *empirical* need to generate realistic simulations, and the *experimental* requirement for pre-selecting network metrics according to the experimental design, and maintaining analytic tractability.

One should note, however, that even with these improvements, the flexibility of the procedure might generate networks that rarely exist in real social structures. Also, since our focus is on global topological metrics, there might be a gap between local and global properties of the network (e.g. high local clustering in one of the bins, but an overall low clustering). Thus, when applying this procedure to simulate real life scenarios, one should carefully choose the range of values for the metrics, and eliminate combinations of values that do not represent the type of networks in interest.

To increase the generalizability of our results in real scenarios, an additional analysis was conducted on a small sample of real networks. It is described in the result section.

4. An agent-based model for simulating diffusion

An agent-based model is used here to describe the diffusion process, in line with the agent-based econophysics approach [36]. Agent-based models are increasingly used for describing complex economic systems, and have been extensively used to describe diffusion of innovations [37,38]. Here, a network with a pre-specified structure is created through the random-graph-with-a-planted-partition generation process described above, and an agent-based model is used to simulate the process of new product diffusion in this network. The model is composed of 1000 agents, and is used to describe the diffusion of both a monopolistic firm with no competition and a duopoly with two competing firms.

4.1. Adoption probabilities

For each network, we simulate the diffusion of the adoption of a new product. Assume first a monopolistic case. Time is discrete, the market starts with all agents at state "0", and when an agent adopts, its state changes to "1". We assume that adoption is uni-directional, that is, an agent can convert from 0 to 1, but cannot dis-adopt and convert from 1 to 0. The transition rule is based on classical diffusion theory, which suggests that the adoption decision is a result of the combined influence of two factors: *external influence*, represented by the probability δ_i that an agent *i* will be influenced by sales people, advertising, promotions, and other marketing efforts[4]; and *internal influence*, which refers to the influence of all means of social interaction such as word of mouth, or imitation. We denote by q_i the susceptibility of agent i to the internal influence of a single other agent, i.e., the probability that a given agent in a given time period will convince agent i to adopt.

The agent activation rule used here is based on a competing risk, or a cascade, approach, where at time t, each prior adopter connected to agent i independently tries to convince i to adopt. Thus, the discrete-time hazard of i to adopt is 1 minus the probability that all these adopters, as well as the advertising efforts, failed the task: $P_i(t) = 1-(1-\delta_i)(1-q_i)^{S_i(t)}$, where $S_i(t)$ is the number of adopters in i's personal social network [39].

Note that this formulation is not the only means of describing contagion in agent-based models. Other models, such as those based on the Ising model analogue, have used a threshold-based approach, where the agent changes states when a certain threshold of utility is reached [6]. However, the competing-risk formulation is more appropriate for modeling new product diffusion, since it converges to the Bass model as the discrete time interval reduces to zero [40]. In addition, this approach considers the overall influence from all agents connected to the potential adopter i, unlike other models, which choose randomly the agent that influences i [24].

Libai, Muller and Peres [20] extended this monopolistic model to describe adoption in a competitive scenario, as follows: if two firms, A and B, compete in the market, then an agent can be in one of three states "0", "A" or "B". Each of the competing firms is assigned its own values for external influence, δ_{iA} and δ_{iB}, and for the internal influence of a single agent, q_{iA} and q_{iB}. Adopters of A and B each independently influence a potential adopter i to adopt their respective firms. The probability of i being successfully convinced by at least one adopter of A or B is given by:

$$P_i^A(t) = 1-(1-\delta_{iA})(1-q_{iA})^{S_i^A(t)}; \qquad P_i^B(t) = 1-(1-\delta_{iB})(1-q_{iB})^{S_i^B(t)},$$

where S_i^A and S_i^B denote all agents in i's personal social network who have adopted either A or B. Now, in a discrete time period t, there could be one of three scenarios: (a) Agent i is convinced to adopt from A but is not convinced to adopt from B. The probability of this happening is $P_i^A(1-P_i^B)$. (b) Agent i is convinced to adopt from B only. The probability of this is $P_i^B(1-P_i^A)$. (c) Agent i is persuaded to adopt from both A and B, but as per the model assumptions has to choose one of the two. The probability of this happening is $P_i^A P_i^B$, and the agent adopts from A rather than B according to the ratio of the probabilities, $\lambda_{iA} = \frac{p_i^A}{p_i^A+p_i^B}$. The probabilities of i actually adopting from firm A, adopting from firm B, or not adopting from either are, respectively:

$$\begin{aligned} &P_i(adopt\ A) = P_i^A(1-P_i^B) + \lambda_{iA}P_i^A P_i^B \\ &P_i(adopt\ B) = P_i^B(1-P_i^A) + \lambda_{iA}P_i^B P_i^A P_i(adopt\ none) = (1-P_i^A)(1-P_i^B) \end{aligned} \tag{4}$$

where $\lambda_{iA} = \frac{p_i^A}{p_i^A+p_i^B}$, $\lambda_{iB} = 1-\lambda_{iA}$.

In the simulation, for each agent in each period, the adoption probability is realized by drawing a random number from a uniform distribution and comparing it to adoption probabilities $P_i(adopt\ A)$ and $P_i(adopt\ B)$. If this number is between 0 and $P_i(adopt\ A)$, A is adopted; if it is between $P_i(adopt\ A)$ and $P_i(adopt\ A) + P_i(adopt\ B)$, B is adopted; and if it is between $P_i(adopt\ A) + P_i(adopt\ B)$ and 1 there is no adoption (since the number is randomly drawn, the order A, B does not matter).

A single run of the simulation (namely, an adoption process starting from zero with a given network structure, δ and q) ends after 30 time periods, which is consistent with common practice in similar models [40]. Given the parameter values used here, the 30 time periods are such that most of the market has adopted by that time.

For simplicity, δ_{iA} and δ_{iB} are assumed to be equal for firms A and B, and identical across agents. To take into account the possible heterogeneous nature of customer propensity to be affected by others, the value of q is assumed to be normally distributed throughout the network. For robustness, we also examined cases in which q was distributed in a power law distribution with the power-law exponent parameter simulated in the commonly used range of 2–3. We also looked at a uniform distribution in which the range was plus minus the standard deviation used in the Normal distribution analysis. The results reported next are robust to the specification of q.

The values of δ and q were chosen to be consistent with previous research regarding the ranges of these parameter values in diffusion models and in agent-based models (e.g. Refs. [20,40]). The parameter δ was assigned the following values: $\delta = 0.001, 0.005, 0.01, 0.015, 0.02$. The values of q differ across networks: In networks with high average degrees, our preliminary simulations show that the interesting dynamics occur for lower values of q, (since the combination of high q values and high degree generates almost instantaneous diffusion). Therefore, two sets of q values were used: For the 80 networks with lower average degrees, we used a distribution with an average q of 0.04, 0.08, 0.1, 0.12, 0.16, and for the 80 networks with higher average degrees, we used distributions with an average q of 0.005, 0.01, 0.02, 0.03, 0.04. The experimental design was a full factorial experiment, where for each network we tested all combinations of the different values of δ and q (from the appropriate set). Thus, the overall experimental design was LHS + full factorial: LHS for the network parameters, and full factorial for the diffusion parameters.

4 In the diffusion literature δ is often denoted as p. Here, p is used to represent probabilities.

For each of the 160 networks, we ran the simulation 6000 times: for each combination of δ and q as elaborated above, for both monopolistic and duopoly scenarios, and, to control for random effects, for each parameter set we ran 120 simulations. Simulations were conducted using C++ code. The overall runtime on an Intel 3.1 GHz i5-2400 core processor, 16 GB RAM was 47 days.

4.2. The time value of diffusion

An effective diffusion process is quick, and it affects a large number of network members. To measure the effectiveness of diffusion, we evaluated the NPV of the number of adopters, that is, the discounted sum of the number of adopters. Thus, for a given firm μ the NPV is $\sum_{t=1}^{T} S_{\mu}(t)/(1+d)^t$, where $S_{\mu}(t)$ is the number of agents who adopted from firm μ ($\mu = A, B$) at time period t. Note that in our simulations, the firms are symmetric, so $S_A(t) = S_B(t)$. T is the total time horizon, and d is the discount factor. In the simulation, a standard discount factor of 10% is used.

5. Results

5.1. Core result

To estimate the impact of network topology on diffusion, we examined how the NPV of the number of adopters is dependent on the various topological metrics and diffusion parameters. Fig. 5 comprises three graphs, illustrating, respectively, the dependence of the ln *_NPV* on each of the three topological metrics we used. For illustration purposes we present values for $\delta = 0.005$ and $q = 0.02$, averaging the values of each remaining variable across all simulation runs and across all the networks. The ln was used to enable a comparison to be made between the coefficients of the monopoly and duopoly cases.

To evaluate the relative role of each of the topological metrics in the diffusion process, we ran a multivariate regression, where ln *_NPV* was regressed simultaneously against the topological metrics and diffusion parameters. The regression data were pooled over the 160 networks. The explanatory variables were the average degree (*Degree*), the relative degree of hubs (*Rel_hubs*), and the clustering coefficient (*Cluster*). The average path length (*Av_path*) was measured and used as a control variable. The diffusion parameters δ and q were mean-centralized. The resultant regression equation was the following:

$$\ln\_NPV = \alpha_0 + \alpha_1 Degree + \alpha_2 Rel\_hubs + \alpha_3 Cluster + \alpha_4 Av\_path + \alpha_5 \delta + \alpha_6 q + \varepsilon. \quad (5)$$

For data smoothing, and to enable OLS regression, each data point in the regression was the average over the 120 runs with the same parameter combination.[5] The estimation was performed separately for the monopoly and duopoly conditions. For the duopoly case the NPV of one of the firms (firm A) was used; since both firms are identical, the choice of A or B is equivalent.

The correlation table for the explanatory variables is displayed in Table 1. The rightmost column of the table presents the Variance Inflation Factor (VIF), a standard measure for multicollinearity. Since the experimental design is a full factorial for the diffusion parameters + Latin hypercube sampling for the networks, we expect all correlations to be small, and VIF < 10 for all variables. Note the high correlation (although still small VIF) between the internal influence parameter q, and *Degree*. Recall that as described above, two sets of q were chosen, according to the average degree. To verify that our results are not biased by this multicollinearity, we performed the analysis separately for each set of 80 networks, and the results show the same pattern.

The results are displayed in Table 2. The table presents the regression coefficients for the monopoly and duopoly cases. In both cases, the relative degree of hubs and the average degree have a positive and significant influence. That is, the higher is the average degree, and the higher is ratio between the average degree of the top 10% most connected agents and the overall average degree, the more effective the diffusion process. This result contributes to the controversy as to the importance of social hubs to the diffusion process: When controlling for diffusion parameters and other network metrics, the impact of the average degree of social hubs is positive. Note, that these results cannot indicate on the relative strength of the two topological metrics, since their relative impact depends on the estimation method (averaged or not averaged over repetitions), and, as we see below, changes across datasets and asymmetry conditions.

The effect of the clustering coefficient is negative, indicating that global clustering has a negative effect on diffusion. This finding contributes to the discussion on the role of transitivity in network processes [11]: While transitivity might have benefits, in the context of diffusion its impact is negative. By varying the clustering coefficient independently of other network metrics, and testing it simultaneously with the other metrics, we have managed to isolate its effect and assess its relative role in diffusion.

[5] Alternatively, the parameters can be estimated without the averaging, through a mixed random effect model: $\ln\_NPV_{ijk} = \alpha_0 + \alpha_1 Degree + \alpha_2 Rel\_hubs + \alpha_3 Cluster + \alpha_4 Av\_path + \alpha_5\delta + \alpha_6 q + \gamma_{ij} + \varepsilon_{ijk}$, where i is the network, j is the δ–q combination, and k is the repetition, $k = 1$–120. Number of observations is 480,000. The estimation uses maximum likelihood, and the resulting parameter estimates are: for monopoly ($\alpha_0 = 5.16,\ \alpha_1 = 0.04,\ \alpha_2 = 0.04,\ \alpha_3 = -0.95,\ \alpha_4 = -0.003,\ \alpha_5 = 100.8,\ \alpha_5 = 4.26$); for duopoly: ($\alpha_0 = 4.59,\ \alpha_1 = 0.08,\ \alpha_2 = 0.02,\ \alpha_3 = -2.87,\ \alpha_4 = 116.79,\ \alpha_5 = 4.42$). Both models converge, all parameters are significant with $p < 0.0001$. These estimations are comparable with those of Table 2. Note the differences in estimation with respect to $\delta(\alpha_5)$. These are expected, since averaging over repetitions might lead to underestimation of parameters [41,42].

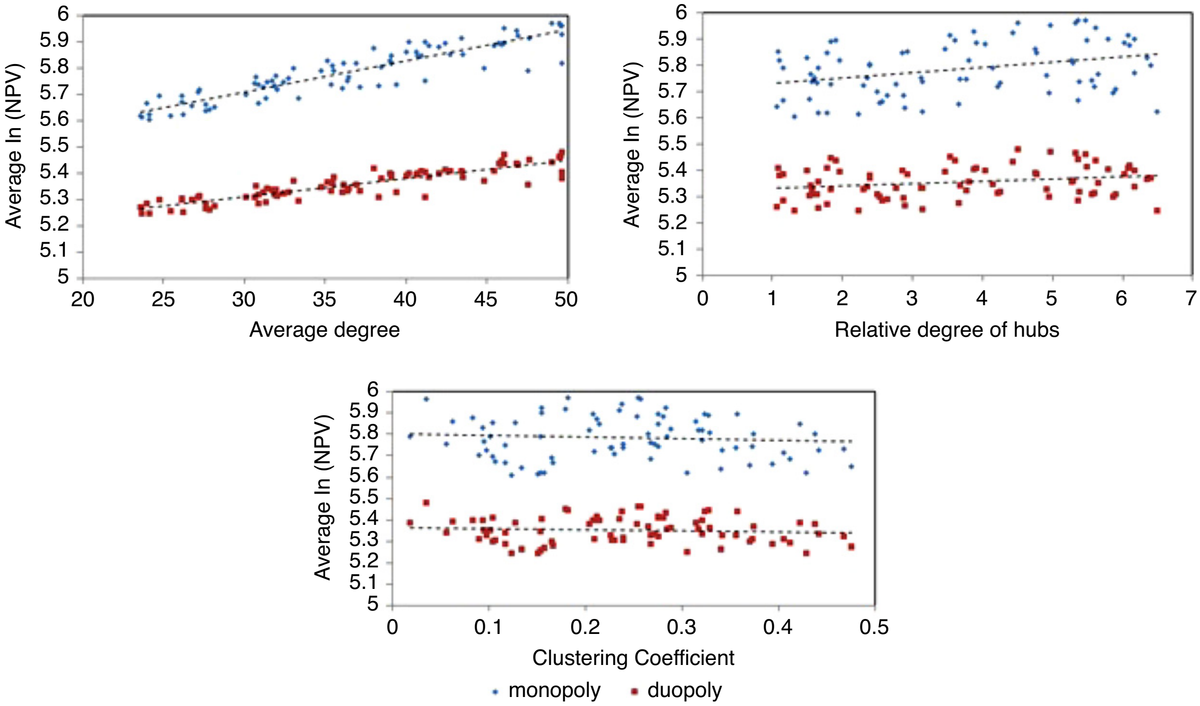

Fig. 5. Network performance as a function of topological metrics, for monopoly and duopoly. $\delta = 0.005$, $q = 0.02$, averaged on all other parameters and on all networks.

Table 1
Correlation table and multicollinearity index.

	δ	q	Degree	Rel_hubs	Cluster	VIF
δ	1					1
q	0	1				1.95
Average degree	0	−0.70	1			2.1
Relative degree of hubs	0	0.14	−0.17	1		1.15
Clustering coefficient	0	−0.11	0.17	0.29	1	1.15
Average path length	0	0.12	−0.22	0.05	−0.06	1.05

All correlations are significant at $p < 0.0001$.

Table 2
Estimation results in monopolistic and duopolistic markets.

	ln(NPV)monopoly	Standardized monopoly**	ln(NPV)duopoly for firm *A*	Standardized duopoly**
Intercept (α_0)	5.43 (0.018)*	5.97*	4.99 (0.015)*	5.42*
Average degree (α_1)	0.02 (0.000)*	0.23*	0.01 (0.000)*	0.19*
Relative degree of hubs (α_2)	0.05 (0.003)*	0.08*	0.03 (0.002)*	0.06*
Clustering coefficient (α_3)	−0.17 (0.041)*	−0.02*	−0.13 (0.03)*	−0.016*
Average path length (α_4)	−0.007 (0.000)*	−0.1*	−0.005 (0.000)*	−0.08*
δ (α_5)	28.6 (0.69)*	0.19*	27.61 (0.58)*	0.18*
q (α_6)	6.25 (6.26)*	0.30*	4.98 (0.11)*	0.24*
Adj R^2	0.54	0.54	0.55	0.55
N (no of observations)	4000	4000	4000	4000

* p value < 0.0001; Standard error is brackets.
** Standard errors for all variables are ∼0.005.

The effect of δ and q is positive and significant, with δ having a considerably large effect. The average path length, used here as a control variable, has a small and significant negative effect. The larger the average distance between two nodes, the slower is diffusion.

Interestingly, these results are robust to the competitive market structure and are the same for both monopolistic and duopolistic markets. All results are significant with $p < 0.0001$.

Table 3
Correlations, descriptive statistics and estimation for 12 real-life connectivities.

(a) Descriptive statistics	Mean	Standard deviation	Minimum	Maximum
δ	0.01	0.007	0.001	0.02
q	0.105	0.06	0.001	0.32
Average degree	9.49	10.33	2.6	42.3
Relative degree of hubs	4.92	1.8	2.31	7.3
Clustering coefficient	0.22	0.14	0.05	0.5

(b) Correlations + multicollinearity index[*]	δ	q	Degree/size	Rel_hubs[a]	Cluster/size	VIF
δ	1					1
q	0	1				2.1
Average degree/size	0	0.13	1			4.43
Relative degree of hubs[a]	0	−0.13	−0.68	1		2.65
Clustering coefficient/size	0	0.62	0.64	−0.43	1	3.54

(c) Estimation results	ln(NPV)monopoly	Standardized monopoly[***]	ln(NPV)duopoly for firm *A*	Standardized duopoly[***]
Intercept (α_0)	6.22 (0.121)[**]	6.85[**]	5.71 (0.12)[**]	6.28[**]
Average degree/size (α_1)	2.63 (3.45)	0.04	0.32 (3.43)	0.005
Relative degree of hubs (α_2)	0.25 (0.02)[**]	0.45[**]	0.23 (0.02)[**]	0.42[**]
Clustering coefficient/size (α_3)	−4281 (245)[**]	−0.88[**]	−4069 (244)[**]	−0.83[**]
δ (α_4)	23.15 (3.92)[**]	0.15[**]	23.58 (3.90)[**]	0.16[**]
q (α_5)	5.51 (0.63)[**]	0.33[**]	4.7 (0.62)[**]	0.29[**]
Adj R^2	0.82	0.82	0.81	0.81
N (no of observations)	300	300	300	300

[*] p value on all correlations < 0.05.
[**] p value < 0.0001; Standard error in brackets.
[***] Standard errors are 0.03–0.05.
[a] Since the relative degree of hubs is defined as the ratio between hub degree and overall degree, normalization of both by size leaves the variable unchanged.

5.2. Additional test with real networks

To show that the above results might apply for real networks, we ran the diffusion simulations described above using real-life connectivity data on the 12 social network structures described in Ref. [20]. Explanatory variables for each network were normalized by network size (number of nodes), δ and q were mean-centralized. Due to high correlations and small sample, average path length is not part of this model. Thus, the regression model is

$$\ln\_NPV = \alpha_0 + \alpha_1 Degree + \alpha_2 Rel\_hubs + \alpha_3 Cluster + \alpha_4\delta + \alpha_5 q + \varepsilon. \tag{6}$$

Table 3 presents the correlations, descriptive statistics, and estimation results. Note, that although the VIF values are satisfactorily small (<10), correlations among the explanatory variables are high. As with the simulated connectivity, the clustering coefficient has a negative impact and the average degree and relative degree of hubs have a positive impact. The average degree effect is insignificant.[6]

The parameter estimations of Table 3 should be interpreted with caution. Sample size is small (12 networks only) and is not representative of the social network universe. Also, explanatory variables are correlated, raising concern for multicollinearity, despite the low VIF. This analysis aims to demonstrate some external validity of our results. Further validation of our insights on real networks is required as part of future research.

5.3. Marketing efforts and network structure

An intriguing question resulting from this analysis with respect to the duopoly scenario relates to the interactions between marketing efforts and topological metrics. Can investment of marketing efforts by one of the firms moderate or enhance the effect of topological metrics? This is especially interesting with respect to the relative degree of hubs. Can the dependence of performance of the relative degree of hubs be moderated or enhanced by investing high marketing efforts? Fig. 6 illustrates this graphically—showing the dependence of the $\ln\_NPV$ of firm A on the relative degree of hubs, in the duopoly scenario, for $\delta_A/\delta_B = 1$, and $\delta_A/\delta_B = 4$. For illustration purposes we present values for $\delta_B = 0.01$and $q = 0.02$,

[6] Using a mixed random effect model on the raw data (without averaging) leads to: for monopoly ($\alpha_0 = 6.6,\ \alpha_1 = 3.92,\ \alpha_2 = 0.22,\ \alpha_3 = -5762,\ \alpha_4 = 39.92,\ \alpha_5 = 4.99$); for duopoly: ($\alpha_0 = 6.15,\ \alpha_1 = 8.94,\ \alpha_2 = 0.18,\ \alpha_3 = -5952,\ \alpha_4 = 40.4,\ \alpha_5 = 4.07$). Both models converge, all parameters are significant with $p < 0.001$. Results are comparable to those of Table 3, with the typical underestimation occurring when averaging over repetitions.

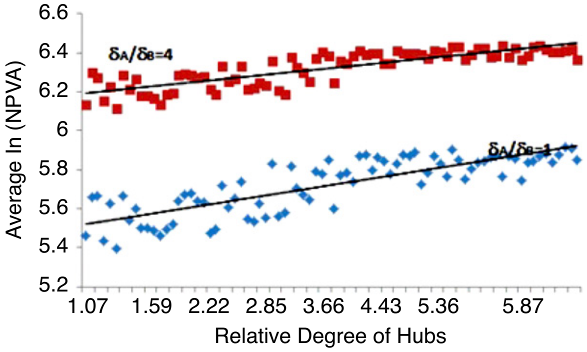

Fig. 6. Network performance as a function of the relative degree of hubs, for asymmetric duopoly, for $\delta_A/\delta_B = 1$ and $\delta_A/\delta_B = 4$. $\delta_B = 0.01$, $q = 0.02$, averaged on all other parameters and on all networks.

Table 4
Estimation results in duopoly, where firms differ in the value of δ.

	ln(NPVA)	ln(NPVA) standardized
Intercept (α_0)	5.27 (0.01)*	5.48(0.009)*
Average degree (α_1)	0.005 (0.000)*	0.08 (0.003)*
Relative degree of hubs (α_2)	0.05 (0.001)*	0.09 (0.003)*
Clustering coefficient (α_3)	−0.1 (0.03)*	−0.01 (0.003)*
Relative strength δ_A/δ_B (α_4)	0.21 (0.003)*	0.21 (0.003)*
Average path length (α_5)	−0.004 (0.000)*	−0.05 (0.001)*
δ_B (α_6)	16.7 (0.18)*	0.11 (0.001)*
q (α_7)	1.65 (0.03)*	0.08 (0.001)*
Rel_Strength* Degree (α_8)	−0.0003 (0.0000)*	−0.01 (0.004)*
Rel_Strength* Rel_hubs (α_9)	−0.004 (0.0000)*	−0.03 (0.005)*
Rel_Strength* Cluster (α_{10})	−0.009 (0.01)	−0.003 (0.004)
Adj R^2	0.68	0.68
N (no of observations)	20 000	20 000

* p value < 0.001; Standard error in brackets.

averaging the values of each remaining variable across all simulation runs and across all the networks. As the figure indicates, the two trend lines are practically parallel, implying a negative answer to this question.

To formally test the effect of interactions, we re-ran the simulations for duopoly in asymmetric conditions, where $\delta_B = 0.001, 0.005, 0.01, 0.015, 0.02$, and δ_A/δ_B (*Rel_Strength*) $= 1, 1.5, 2, 3, 4$, in a full factorial scenario. Namely, an experimental setting of 160 networks $\times$ 5 values of q $\times$ 5 values of δ_B $\times$ 5 values of δ_A/δ_B. To save running time, each combination was run for 60 repetitions instead of 120. Each data point in the regression was the average over the 60 runs with the same parameter combination.

To capture the interactions between the relative strength of external influence and the topological metrics we ran the following model:

$$\ln\_NPV = \alpha_0 + \alpha_1 Degree + \alpha_2 Rel\_hubs + \alpha_3 Cluster + \alpha_4 Rel\_Strength + \alpha_5 Av\_path + \alpha_6 \delta_B + \alpha_7 q + \alpha_8 Rel\_Strength * Degree + \alpha_9 Rel\_Strength * Rel\_hubs + \alpha_{10} Rel\_Strength * Cluster + \varepsilon. \quad (7)$$

For the regression, δ_B and q were mean-centralized. Table 4 displays the estimation results. While the main effects do not change, the interactions are weak or insignificant. Similar results are obtained when running the mixed random effect model on the raw data without averaging. This result indicates that the relationship between topological metrics and diffusion performance of a firm do not depend on the firm's relative level of marketing efforts. Changes in δ_A/δ_B do not moderate or enhance the extent to which the three topological metrics influence the NPV of A. Of course, a higher δ_A influences A's NPV through the main effect, but it does not change the functional dependence between the NPV and the network topology.

6. Discussion and conclusions

This paper focuses on the effects of average degree, relative degree of hubs and clustering coefficient on the diffusion of a new product. We introduce a network generation procedure, based on random graphs with a planted partition, to generate and experimental design of 160 networks encompassing a large range of values for the parameters under investigation. Using agent-based models, we simulate diffusion processes on these networks, for monopolistic and duopolistic markets.

By directly manipulating the topological metrics and measuring their impact simultaneously, we can *isolate the relative influence* of each metric. We found that the relative degree of hubs, as well as the average degree, have strong and positive effects on diffusion. The clustering coefficient, however, has a negative impact on diffusion: the higher the level of global clustering, the weaker the diffusion process.

These findings shed light on the ways in which underlying network topologies influence diffusion, and they contribute to two ongoing discussions in the network literature. The first result emphasizes the importance of hubs, i.e., nodes whose degrees are relatively high compared with the degrees of the rest of the population. Specifically, our analysis suggests that a degree distribution that tends towards a power-law is likely to be associated with a more effective diffusion process (in terms of NPV) compared with a more uniform degree distribution. This result demonstrates that the role of social hubs in diffusion processes is intricate and depends on the aspect of diffusion that is being tested: While the relative degree of hubs might be less important when measuring the length of individual cascades of influence (the length of a single information, or influence chain) [9], it is important to the overall effectiveness of diffusion.

Our finding regarding the negative impact of clustering contributes two interesting insights to the ongoing discussion on the pros and cons of network transitivity [11]: First, the result sheds light on the role of transitivity, in isolation from other network metrics. While most other studies have compared network types, thereby considering clustering in combination with other network metrics, here we measure the direct effect of clustering on diffusion. Second, our result demonstrates the drawbacks of transitivity in the context of diffusion. It seems that the redundancies generated by high clustering impede diffusion.

Methodologically, we introduce here a method for network generation, which enables to generate networks with pre-specified average degree, relative degree of hubs, and clustering coefficients. This is a variation on the l-partition graph discussed in graph partitioning and community detection problem. The variation we suggest overcomes some of the limitations of the pure l-partition graphs, to create more realistic networks. It is also unique since it enables to vary independently the degree distribution as well as the clustering. This method could be further extended to pre-specify other network dimensions such as path length or degree of separation. Moreover, this method enables generating hierarchical structures such as communities within communities, through a manipulation of the bin sizes and connection probabilities.

Note that although this paper studies a wide range of parameter values, its findings cannot automatically be generalized to all network types. Since the network creation procedure is based on random graphs, the degree distribution is a Poissonian mix function, and is different from some degree distributions found in real networks, such as the power law distribution. The results are not expected to be fundamentally different for such networks; however, further research is needed to verify this empirically.

Acknowledgments

This paper was supported by the Israel Science Foundation, the Israeli Association for Internet Study, and the Kmart fund. My thanks to Michael Krivelevich for discussion on random graphs with a planted partition. I also thank Meir Karlinsky for his insightful remarks on sampling methods. I thank Zeev Rudnick, Barak Libai and Eitan Muller for their comments.

References

[1] F.M. Bass, A new product growth model for consumer durables, Manag. Sci. 15 (1969) 215–227.
[2] J. Goldenberg, B. Libai, E. Muller, Talk of the network: a complex systems look at the underlying process of word-of-mouth, Mark. Lett. 12 (2001) 211–223.
[3] M. Barthelemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Velocity and hierarchical spread of epidemic outbreaks in scale-free networks, Phys. Rev. Lett. 92 (2004) 178701-1-4.
[4] M. Granovetter, The strength of weak ties, Am. J. Sociol. 78 (1973) 1360–1380.
[5] J. Goldenberg, S. Han, D. Lehmann, J. Hong, The role of hubs in the adoption process, J. Mark. 73 (2009) 1–13.
[6] C.E. Laciana, S.L. Rovere, Ising-like agent-based technology diffusion model: adoption patterns vs. seeding strategies, Physica A 390 (2011) 1139–1149.
[7] F. Stonedahl, W. Rand, U. Wilensky, Evolving viral marketing strategies, in: Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation, ACM, New York, 2010, pp. 1195–1202.
[8] M. Uchida, S. Shirayama, Influence of a network structure on the network effect in the communication service market, Physica A 387 (2008) 5303–5310.
[9] D.J. Watts, P.S. Dodds, Influentials, networks, and public opinion formation, J. Consum. Res. 34 (2007) 441–458.
[10] K. Keller, J. Berry, The Influentials: One American in Ten Tells the Other Nine How to Vote, Where to Eat, and What to Buy, The Free Press, New York, 2003.
[11] M.J. Burger, V. Buskens, Social context and network formation: an experimental study, Soc. Networks 31 (2009) 63–75.
[12] J. Leskovec, L.A. Adamic, B.A. Huberman, The dynamics of viral marketing, ACM Trans. Web 1 (2007) 5–51.
[13] R. Albert, A.L. Barabási, Statistical mechanics of complex networks, Rev. Modern Phys. 74 (2002) 47–97.
[14] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440–442.
[15] M.E.J. Newman, D.J. Watts, Scaling and percolation in the small-world network model, Phys. Rev. E 60 (6) (1999) 7332–7342.
[16] M.E.J. Newman, D.J. Watts, S.H. Strogatz, Random graph models of social networks, Proc. Natl. Acad. Sci. USA 99 (Suppl. 1) (2002) 2566–2572.
[17] P. Erdős, A. Rényi, On random graphs, Publ. Math. Debrecen 6 (1959) 290–297.
[18] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, Complex networks: structure and dynamics, Phys. Rep. 424 (2006) 175–308.
[19] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256.
[20] B. Libai, E. Muller, R. Peres, Decomposing the value of word-of-mouth seeding programs: acceleration vs. expansion, J. Mark. Res. 50 (2013) 161–176.
[21] M. Faloutsos, P. Faloutsos, C. Faloutsos, On power-law relationships of the Internet topology, ACM SIGCOMM Comput. Commun. Rev. 29 (1999) 251–262.
[22] M.A. Janssen, W. Jager, Simulating market dynamics: interactions between consumer psychology and social networks, Artif. Life 9 (2003) 343–356.
[23] L. Deng, Y. Liu, F. Xiong, An opinion diffusion model with clustered early adopters, Physica A 392 (2013) 3546–3554.
[24] A. Barrat, M. Weigt, On the properties of small-world network models, Eur. Phys. J. B 13 (2000) 547–560.
[25] L.da F. Costa, F.A. Rodrigues, G. Travieso, P.R. Villas Boas, Characterization of complex networks: a survey of measurements, Adv. Phys. 56 (2007) 167–242.

[26] S. Fortunato, Community detection in graphs, Phys. Rep. 486 (2010) 75–174.
[27] A. Condon, R.M. Karp, Algorithms for graph partitioning on the planted partition model, Random Struct. Algorithms 18 (2001) 116–140.
[28] J.C. Helton, F.J. Davis, Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems, Reliab. Eng. Syst. Saf. 81 (2003) 23–69.
[29] M.D. McKay, R.J. Beckman, W.J. Conover, Comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics 21 (1979) 239–245.
[30] R.H. Myers, D.C. Montgomery, C.M. Anderson-Cook, Response Surface Methodology: Process and Product Optimization Using Designed Experiments, third ed., Wiley, New York, 2009.
[31] L.P. Swiler, R. Slepoy, A.A. Giunta, Evaluation of sampling methods in constructing response surface approximations, Paper AIAA-2006-1827 in Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Newport, Rhode Island, May 1–4, 2006.
[32] M.E.J. Newman, Random graphs as models of networks, Handb. Graphs Netw. (2003) 35–68.
[33] M.O. Jackson, B.W. Rogers, Meeting strangers and friends of friends: how random are social networks? Amer. Econ. Rev. (2007) 890–915.
[34] B. Karrer, M.E.J. Newman, Stochastic blockmodels and community structure in networks, Phys. Rev. E 83 (2011) 016107.
[35] A. Lancichinetti, S. Fortunato, Community detection algorithms: a comparative analysis, Phys. Rev. E 80 (2009) 056117.
[36] C. Schinckus, Between complexity of modelling and modelling of complexity: an essay on econophysics, Physica A 392 (2013) 3654–3665.
[37] M. Haenlein, B. Libai, Targeting revenue leaders for new products, J. Mark. 77 (2013) 1–55.
[38] R. Garcia, Uses of agent-based modeling in innovation/new product development research, J. Prod. Innovat. Manag. 22 (2005) 380–398.
[39] J. Goldenberg, B. Libai, E. Muller, Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata, Acad. Mark. Sci. Rev. 9 (2001) 1–18.
[40] J. Goldenberg, O. Lowengart, D. Shapira, Zooming in: self emergence movements in new-product growth, Mark. Sci. 28 (2) (2009) 274–292.
[41] F.B. Hu, J. Goldberg, D. Hedeker, B.R. Flay, M.A. Pentz, Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes, Am. J. Epidemiol. 147 (1998) 694–703.
[42] J.M. Neuhaus, J.D. Kalbfleisch, W.W. Hauck, A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data, Int. Statist. Rev. (1991) 25–35.