



Finding early adopters of innovation in social networks

Balázs R. Sziklai^{1,2} · Balázs Lengyel^{1,3}

Received: 24 October 2022 / Revised: 26 November 2022 / Accepted: 28 November 2022 / Published online: 8 December 2022
© The Author(s) 2022

Abstract

Social networks play a fundamental role in the diffusion of innovation through peers' influence on adoption. Thus, network position including a wide range of network centrality measures has been used to describe individuals' affinity to adopt an innovation and their ability to propagate diffusion. Yet, social networks are assortative in terms of susceptibility and influence and in terms of network centralities as well. This makes the identification of influencers difficult especially since susceptibility and centrality do not always go hand in hand. Here, we propose the Top Candidate algorithm, an expert recommendation method, to rank individuals based on their perceived expertise, which resonates well with the assortative mixing of innovators and early adopters in networks. Leveraging adoption data from two online social networks that are assortative in terms of adoption but represent different levels of assortativity of network centralities, we demonstrate that the Top Candidate ranking is more efficient in capturing innovators and early adopters than other widely used indices. Top Candidate nodes adopt earlier and have higher reach among innovators, early adopters and early majority than nodes highlighted by other methods. These results suggest that the Top Candidate method can identify good seeds for influence maximization campaigns on social networks.

Keywords Online social networks · Innovation adoption · Network centrality measures · Top Candidate ranking · Homophily

1 Introduction

Most individuals adopt an innovation by imitating their influential peers (Rogers 1962; Bass 1969) that underlines the role of social networks in the diffusion of new products, technologies or ideas (Granovetter 1978; Valente 1996). Network scientists argue that the structure of social networks can explain the underlying mechanisms of social influence and adoption: highly connected nodes have more influence than others (Pastor-Satorras et al. 2015), while diffusion is

more likely in tightly connected cliques and less likely across them (Centola and Macy 2007). Wang et al. (2019) paint a more nuanced picture as they found that network hubs are effective in spreading simple messages, less connected nodes gain importance in the diffusion of complex stories.

A central part of this discussion has led to the “influence maximization” problem (IM) (Kempe et al. 2003), which aims to identify the ideal seed nodes that a marketing campaign should target to achieve maximum impact, given pre-defined diffusion models. The IM is NP-hard; thus, many use heuristics to find the seed nodes and start optimization by assuming that nodes with high network centrality (e.g., degree) are influential spreaders (Kitsak et al. 2010; De Arruda et al. 2014) and run diffusion simulations, most notably using the linear threshold and the independent cascade models. However, these models fail to capture an important feature that is observed in real-life networks: homophily, the tendency that similar individuals are more likely to be connected than dissimilar ones.

Homophily, also referred to as assortativity in relation to social networks (Newman 2002), is a general phenomenon (McPherson et al. 2001; Cho et al. 2012) that has a

✉ Balázs R. Sziklai
sziklai.balazs@krtk.hu

Balázs Lengyel
lengyel.balazs@krtk.hu

¹ Institute of Economics, Centre for Economic and Regional Studies, Tóth Kálmán u. 4., Budapest 1097, Hungary

² Department of Operations Research and Actuarial Sciences, Corvinus University of Budapest, Fővám tér 8., Budapest 1093, Hungary

³ Corvinus Institute for Advanced Studies, Corvinus University of Budapest, Fővám tér 8., Budapest 1093, Hungary

fundamental role in innovation spreading (Anwar et al. 2021). Ignoring this effect poses a major problem for seed identification in influence maximization when the sole source of information is the network structure (Aral and Dhillon 2018).

Although there are papers that address the role of homophily in network diffusion and papers that consider innovators and early adopters in influence spreading (see Sect. 2 for an overview), none focus on both problems simultaneously. In particular, our paper is the first that aims to identify innovators and early adopters, while taking their assortative mixing into account, with the aim to provide heuristics for seed selection in influence maximization.

This research niche is important since social influence and centrality are difficult to disentangle without knowing at least some of the early adopters of the specific innovation (Banerjee et al. 2013) or assuming homophily in terms of adoption in the network (Toole et al. 2012). Furthermore, central individuals may be reluctant to participate in a campaign or may not be susceptible to the marketing message due to risk-averseness. Subscribing to a new trend or technology needs commitment and entails social risk—not everyone is willing to do that—yet, homophily in terms of risk-taking behavior is a prerequisite of diffusion cascades (Watts 2002). Central agents with many friends may particularly feel the social pressure to be conformist and to avoid eccentric behavior. Innovators and early adopters, on the other hand, are known to possess psychological traits that make them perfect subject for the early market of an innovation (Rogers 2003).

In this paper, we aim to contribute to the above discussion in two ways. First, empirical data on adoption dynamics from two online social networks enable us to investigate how network structure can be useful to identify innovators and early adopters in innovation diffusion. Second, we propose a ranking of the users based on the so-called Top Candidate method (Sziklai 2018)—an expert selection algorithm that exhibits features resembling assortativity.

We compare the Top Candidate ranking with seven well-known centrality measures on two online social networks: iWiW from Hungary and Pokec from Slovakia. Registration days of users are known in both networks, both are assortative in terms of adoption time but represent different levels of assortativity in network centralities. We look at the top 1000 nodes of the Top Candidate ranking and the other seven alternative measures and plot how the date of registration is distributed over time.

We find that the Top Candidate ranking is more efficient in capturing innovators and early adopters than other widely used indicators. Top Candidate nodes adopt earlier and have higher reach among innovators, early adopters and early majority than nodes highlighted by other methods. These results suggest that the Top Candidate method can identify

good seeds for influence maximization campaigns on social networks.

2 Literature overview

2.1 Early adopters as opinion leaders

The identification of innovators and early adopters is key for marketing campaigns and their characterization received considerable attention. The literature converges toward the conclusion that innovators and early adopters stand out from their peers.

Rogers (2003) describes innovators as venturesome individuals who can cope with a high degree of uncertainty, and early adopters as a group with high socioeconomic status. Moore (2014) depicts innovators as technology enthusiasts, or geeks and early adopters as visionaries who are willing to take high risk. In the literature, innovators and early adopters are often grouped together under the term early market (Muller and Yogeve 2006; Moore 2014). Although we do not follow this convention here, we do assume that both groups are susceptible to marketing messages, hence they are good candidates as seeds for influence maximization.

A field study by Brancheau and Wetherbe (1990) supports hypotheses that early adopters were more highly educated, more attuned to mass media, more involved in interpersonal communication, and more likely to be opinion leaders. Eastlick and Lotz (1999) reports that social risk negatively relates to the tendency to be a potential innovator and potential innovators possessed significantly stronger opinion leadership. A dutch survey shows that early adopters are likely to be highly mobile, have a high socioeconomic status, high levels of education and high personal incomes (Zijlstra et al. 2020). Gender imbalance can be also observable for certain products. Plötz et al. (2014) report that early adopters for electric vehicles are predominantly middle-age men. Finally, Muller and Yogeve (2006) provides empirical evidence that the average time at which the main market outnumbers the early market is indeed when 16% of the market has already adopted the product—giving support Rogers (1962)'s division of adopter sets.

Another important concept is market mavenness (Feick and Price 1987). Market mavens are consumers who are highly involved in a market. They have information about many kinds of products and shops, and they enjoy sharing their knowledge. Peers often seek out their opinion and rely on their expertise. Goldsmith et al. (2003) finds that consumer innovativeness and market mavenism positively correlate, although they argue that market mavens and innovators are distinct groups. Nevertheless, market mavens can convince their community and thus their social interaction

is decisive for innovation diffusion, as it was demonstrated in the case of electric vehicles (Seebauer 2015).

Directly related to the context of this study, Lynn et al. (2011) explores the relationship between personality traits of early adopters of social network sites. They report that extraversion, openness and conscientiousness impact positively and significantly on information sharing, and negatively on rumor sharing. On the other hand both, information sharing and rumor sharing impact positively and significantly on the centrality of early adopters. The seemingly contradictory observations can be explained away by separating the social status of opinion leadership and the influencing capacity of the agent which relates more to network centrality.

To sum up, innovators and early adopters stand out in their personal characteristics. Thus, marketing campaigns have usually targeted and labeled them as opinion leaders to convince society. However, both Lynn et al. (2011) and Dedehayir et al. (2017) argue that a distinction has to be made between opinion leadership and innovativeness. Even Rogers (2003) affirms that opinion leaders are not necessarily innovators.

2.2 Early adoption and homophily in network diffusion

In the Influence Maximization framework (Kempe et al. 2003), few papers addressed other node characteristics concentrating in network communities that can help to predict the future popularity of novelty. For example, influential individuals can form clusters that can help the early propagation of an idea (Aral and Walker 2012). Weng et al. (2014) build a predictive model for meme popularity using three classes of features: network topology, community diversity, and growth rate. They found that community related features are the most powerful predictors of future success. Hajdu et al. (2020) study the community structure of public transportation networks and find that transmission probabilities depend on the community structure. Calió and Tagarelli (2021) study attribute-based seed diversification. They argue that a seed set with different characteristics (age, gender, etc.) might be more successful in information-propagation. Rahimkhani et al. (2015) identifies the community structures of the input graph then chooses a number of representative nodes to form the final output of the proposed algorithm.

However, this literature has largely overlooked a phenomenon inherent in social networks and diffusion dynamics alike: the role of homophily (McPherson et al. 2001). It has long been recognized that a behavior can spread in society only when those most prone to it are surrounded by peers who are somewhat less but almost equally open to its adoption (Granovetter 1978). In other words, innovators must be connected to early adopters such that adoption can

penetrate in their communities and later influence the rest of the market too, otherwise the innovation will not spread (Watts 2002). Adoption dynamics can be predicted at small scales only by assuming homophily of adoption (Toole et al. 2012). Despite the importance of adoption homophily in networks, it has been largely ignored in influence maximization modeling (Aral and Dhillon 2018).

Instead, a usual assumption to find the seed nodes for Influence Maximization is that network structure alone can quantify influence. For example, nodes with high network centrality (e.g., degree) are usually considered as influential spreaders (Kitsak et al. 2010; De Arruda et al. 2014).

Finally, the presence of assortativity implies that not every connection is equally important in the diffusion. However, the literature also ignored the problem of determining where the probabilities of influence between users come from (Goyal et al. 2010). Recently, Qiang et al. (2019) proposed two learning models that are aimed at understanding person-to-person influence in information diffusion from historical cascades, while Bóta et al. (2015) and Bóta et al. (2016) considered the Inverse Infection Problem as a way to estimate the hidden edge infection probabilities.

3 Data and methodology

In this paper, we propose the Top Candidate method that can identify innovators and early adopters in social networks more efficiently than other widely used network centrality measures, by using network structure as the only source of information. We compare the ranking induced by the Top Candidate method with seven other centrality measures by using data from two online social networks.

3.1 Data

Our empirical analysis leverages data retrieved from two social media platforms. The first platform is called iWiW (international who is who) that was an early Hungarian version of online social networks aiming to link pre-existing friends and an outstanding online innovation of its time. The iWiW platform existed between 2002 and 2014. It was the most visited website in the country in the mid 2000s, but failed the competition with Facebook that started in Hungary from 2008. Pokec is a still functioning Slovakian dating and chatting website with a purpose of meeting new people.

These data sources provide unique opportunity to understand how network structure can help us identifying early adopters of an innovation. Both data sources contain the date of individual registration to the websites that is used as a proxy of adoption. We define innovators as the first 2.5% and early adopters as the following 13.5% of adopters (Rogers 2003). Data also include the identifiers of friends

that enables us to generate social networks. The iWiW dataset has been used in previous work to describe and model the innovation diffusion process (Török and Kertész 2017; Lengyel et al. 2020; Bokányi et al. 2022).

Here, we use a 10% sample of the iWiW data that contain 271 913 nodes 2 712 587 edges. The Pokec network contains 277 695 nodes and 2 122 778 edges. Access to iWiW data was provided to us by a non-disclosure agreement with the data owner company. Pokec data are open access at <https://snap.stanford.edu/data/soc-pokec.html>.

3.2 The Top Candidate method

Top Candidate (TC) algorithm is a group identification method designed to find experts on recommendation networks (Sziklai 2018, 2021). The algorithm takes a network as an input and outputs a list of experts. With a parameter, $\alpha \in [0, 1]$, we can adjust how exclusive our list should be. Each agent nominates α fraction of their most popular neighbors as experts, where popularity is based on (weighted) in-degree. In the beginning, every agent is labeled as an expert, then in successive rounds, we remove the nominations of agents who were not nominated by anyone until we obtain a stable set. The underlying idea resembles homophily: Experts identify other experts much more effectively than amateurs. Thus, in the set of experts (i) each expert should be nominated by another expert and (ii) each nominee of an expert should be also included in the set—this property is called *stability*. In this paper, we apply this algorithm to identify innovators and early adopters based on the assumption that opinion leaders can be similarly identified in networks like experts.

One advantage of the Top Candidate algorithm is its axiomatic characterization. It is the unique method¹ that satisfies stability, exhaustiveness and decisiveness. Exhaustiveness ensures that all possible experts are recognized on the network, not just a subset, and decisiveness guarantees that at least one expert is selected if reasonable choices are presented. There are other centralities that feature characterizations, most notably PageRank (Wąs and Skibski 2018), Generalized Degree (Csató 2017) and the Shapley value (Shapley 1953; Young 1985), but it is less clear how these relate to socio-demographic properties of the nodes.

3.3 Network centralities

We compute seven other measures on the data to assess their ability in finding innovators and early adopters.

Degree represents the number of connections that a user has. It is a natural benchmark for the user's centrality. Another classical measure is *Harmonic centrality*. It is a distance-based measure proposed by Marchiori and Latora (2000). Harmonic centrality of a node, u , is the sum the reciprocal of distances between u and every other node in the network. Disconnected node pairs have infinite distance, thus the reciprocal is defined as zero. Peripheral agents, who are many handshakes away from most of the other users, thus have a small Harmonic centrality.

PageRank (PR), introduced by Page et al. (1999), is a close relative of Eigenvector centrality (Bonacich 1972). The latter assigns centrality scores to nodes based on the eigenvector of the adjacency matrix of the underlying graph. The method breaks down if the graph is not strongly connected. PageRank rectifies this by (i) Connecting sink nodes (i.e., nodes with no leaving arc) with every other node through a link and (ii) Redistributing some value uniformly among the nodes. Redistribution is parameterised by the so-called damping factor, $\alpha \in (0, 1)$. PageRank was designed to model a random walk on the World Wide Web. We start from an arbitrary webpage. On any subsequent step, we leave the current webpage with equal probability on one of the departing links. After each step, we have a $(1 - \alpha)$ probability to restart the walk at a random node. The probability that we occupy node u as the number of steps tends to infinity is the PageRank value of node u . PageRank composes the core of Google's search engine, but the algorithm is used in a wide variety of applications. The damping value is usually chosen from the interval (0.7, 0.9), here we opted for $\alpha = 0.8$.

Generalized degree discount (GDD) introduced by Wang et al. (2016) was developed specifically for the *independent cascade* network diffusion model. In this model, each active node has a single chance to infect its neighbors, transmission occurring with the probability specified by the arc weights. GDD is a suggested improvement on *Degree Discount* (Chen et al. 2009) which constructs a seed group of size q starting from the empty set and adding nodes one by one using a simple heuristic. It primarily looks at the degree of the nodes but also considers how many of their neighbors are already in the seed set. GDD improves this by also taking into account how many of the neighbors' neighbors are spreaders. The spreading parameter of the algorithm was chosen to be 0.05.

k-core, also referred to as *k-shell*, categorizes nodes into layers (Seidman 1983; Kitsak et al. 2010). First, it successively delete nodes with only one neighbors. These are assigned a *k-core* value of 1. Then, it deletes nodes with two or less neighbors and labels them with a *k-core* value of 2. The process is continued until every node is classified. For

¹ In the Social Choice literature a group identification method, that takes a recommendation network as input and outputs a list of nodes, is called a *collective identity function* CIFs are special kinds of centrality measures that map the node set into $\{0, 1\}^n$ instead of the usual $\mathbb{R}^n$.

Table 1 Group interconnectedness in iWiW. An entry of the matrix shows the portion of links that connects the column group to the row group with respect to the column group's total connections

	Innovators	Early adopters	Early majority	Late majority	Laggards
Innovators	23.4	7.7	2.5	1.3	1.1
Early adopters	39.8	36.4	18.6	11.7	10.5
Early majority	26.8	38.6	48.4	42.5	37.4
Late majority	8.4	14.3	25.2	36.3	35.2
Laggards	1.6	3.0	5.2	8.3	15.9
Total	100.0	100.0	100.0	100.0	100.0

Table 2 Group interconnectedness in Pokec. An entry of the matrix shows the portion of links that connects the column group to the row group with respect to the column group's total connections

	Innovators	Early adopters	Early majority	Late majority	Laggards
Innovators	12.7	4.5	1.6	0.8	0.7
Early adopters	36.0	32.7	16.2	7.3	3.5
Early majority	34.5	45.2	51.7	35.7	15.9
Late majority	13.9	15.6	27.4	45.7	40.9
Laggards	2.9	1.9	3.1	10.5	39.1
Total	100.0	100.0	100.0	100.0	100.0

instance, every node of a path or a star graph is assigned a k -core value of 1, while nodes of a cycle will have a k -core value of 2.

Linear threshold centrality (LTC), as the name suggests, was developed for the linear threshold diffusion model (Riquelme et al. 2018). Given a network, G with node thresholds and arc weights, LTC of a node u represents the fraction of nodes that u and its neighbors would manage to activate as a seed set in the linear threshold model. Since the social networks we used in our analysis had no data on friendship intensity, we decided to assign a uniform unit weight to each connection. Node thresholds was defined as 0.7 times the node degree. That is, a user became activated if 70% of its friends had been active.

Suri and Narahari (2008) define a cooperative game on the network and derive node centrality by computing the Shapley value. In this setting, the Shapley value of a node is the average marginal contribution that a node generates when the seed set is composed by adding nodes one by one and any order of the nodes is equally likely. Every node set is assigned a (characteristic function) value. Marginal contribution of a node u is just the difference between the value of the node set with and without u . There is more than one way how this can be done. We use the G1 game variant proposed by Michalak et al. (2013) who also gave an efficient algorithm to compute the corresponding **Shapley(G1)-value**. In G1, the characteristic function value of a node set C is the number of nodes in C plus the number of neighbors of C . Under this setting, the Shapley value of a node u is calculated as the sum of reciprocals, $\frac{1}{1+\deg(v)}$, for each v belonging to the extended reach of u (the neighbors of u plus u itself).

4 Results

4.1 Homophily of adoption

Before we delve into the performances of centrality measures, let us take a look at the networks themselves. Tables 1 and 2 explore the interconnectedness of adopter groups. iWiW and Pokec paint a similar picture: Typically, there are more connections between subsequent groups in the adoption timeline than between other groups. Innovators are mainly befriended with early adopters, who in turn are mainly connected to early majority and so on.

A number of interesting observations can be made. Firstly, the result reinforces Rogers' classification. It is much more obvious why cascade happens the way it does. Innovators have the biggest impact on early adopters because early adopters are the innovators closest—or at least the most numerous—friends.

Secondly, psychological traits do affect the network structure. Rogers' categorization correlates with risk attitudes, extraversion, openness and a number of other traits. It seems that risk-seeking (extrovert, open-minded, etc.) users prefer the company of other risk-seekers, while risk-averse users are more comfortable with other risk-averse individuals. The results are in line with the findings of Selfhout et al. (2010).

Thirdly, identifying innovators and early adopters does not seem to be a hopeless task anymore. Clearly, these groups form clusters on the network. Thus, there can be centralities that are systematically better in recognizing them.

These observations have a rather remarkable implication. Researchers of influence maximization frequently

Table 3 Assortativity of the iWiW and Pokec networks in terms of network centralities. The assortativity index ranges from -1 to $+1$. Negative values mean that nodes of similar centrality values are not connected, while positive values mean that nodes of similar centrality values are connected

Network centralities	iWiW	Pokec
Degree	0.04	-0.158
Harmonic centrality	0.306	0.137
Page rank	0.02	-0.13
Generalized degree discount	0.02	-0.151
k-shell	0.27	0.227
Linear threshold centrality	0.05	-0.06
Shapley value	-0.004	-0.095
Top Candidate	0.09	-0.083

nodes that belong to innovators or early adopters should be lower in general than thresholds of other nodes. This could be achieved by choosing the thresholds from an interval. Disregarding the underlying structure introduces a systemic bias that may be favorable for some influence maximization algorithms while detrimental to others.

Although the two online social networks are similar in terms of adoption homophily, the assortativity of these networks is different in terms of the network centrality measures described in Sect. 3.3. Both networks are assortative in terms of Harmonic centrality and k-shell measures (Table 3). However, Pokec is disassortative in terms of Degree, Generalized Degree Discount, and PageRank. This means that the identification of innovators and early adopters is carried out on networks in which individuals of similar levels of

Table 4 Overlap in the top 1000 nodes of different centralities on the iWiW network. Measures are ordered according to their distance to degree

	k-core	Harm.	TC	Shapley(G1)	GDD(0.05)	PR	LTC(0.7)	Deg.
k-core	1000	13	39	39	54	49	84	75
Harm.		1000	241	295	290	373	392	374
TC			1000	394	485	477	470	482
Shapley (G1)				1000	652	753	621	629
GDD(0.05)					1000	813	788	828
PR						1000	840	845
LTC(0.7)							1000	935
Deg.								1000

validate their algorithms using simulations with either the linear threshold or the independent cascade diffusion models—these are the most commonly used configurations by far. A basic flaw in these simulations is that thresholds and diffusion probability are chosen at random either independently of the network structure or only having a crude relationship with it. For instance, in the linear threshold model in every simulation, the node thresholds (which signify the tendency for the nodes to adopt an innovation) are generated uniformly at random for each node (Kempe et al. 2003). In the independent cascade model, the two most common propagation setup is the weighted cascade and the trivalency models (Jung et al. 2012). In the first, the propagation probability on each edge equals to the reciprocal of the degree of the source node, while in the latter, it is chosen randomly from the set $\{0.1, 0.01, 0.001\}$.

In light of Tables 1 and 2, these assumptions lead to a highly unrealistic threshold/propagation probability distributions. In order to obtain a realistic network configuration, the distribution should take into consideration the clustering of the adopter sets. For instance, thresholds of

assumed influence are mixed differently.

4.2 Identification of innovators and early adopters

Now we turn to the network centrality indicators and their performances in finding innovators and early adopters. We computed the top 1000 nodes according to eight centrality measures on both iWiW and Pokec. If the 1000th and 1001st node tied under some measure, we discarded nodes of the same centrality value randomly until there were only 1000 nodes in the set.

Tables 4 and 5 show the overlap between the top 1000 nodes of the centralities that we employed in this paper on the iWiW and Pokec networks. Each centrality genuinely differs from the others, although LTC, GDD and PageRank somewhat overlap with Degree on both networks. In general, k-core, TC and Harmonic centrality contain more nodes that are uniquely represented by those centralities.

Table 6 compiles the average and median date of registration for the top 1000 nodes. Centralities are ordered by the

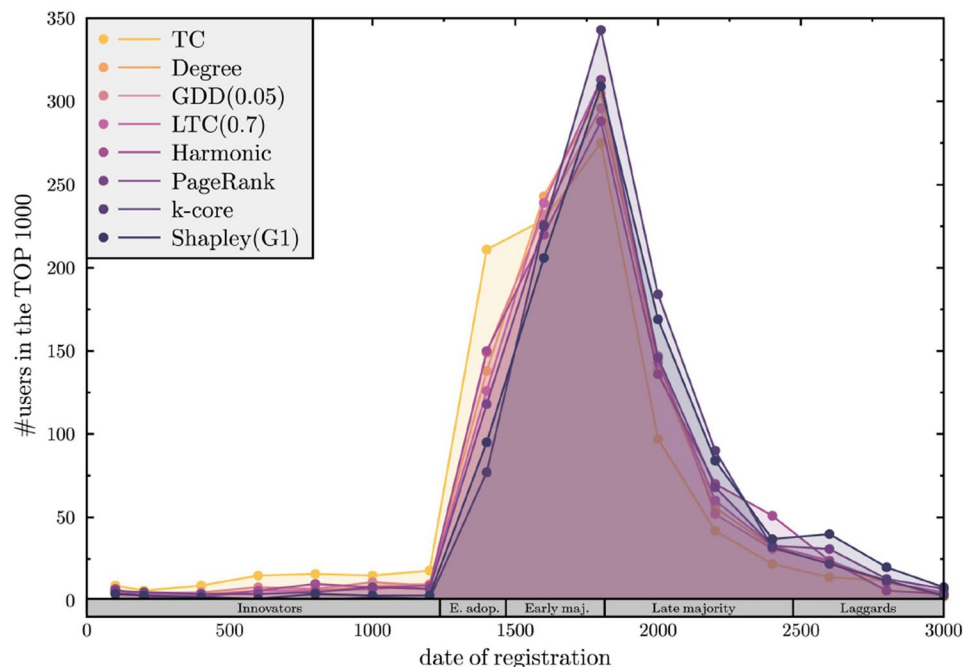
Table 5 Overlap in the top 1000 nodes of different centralities on the Pokec network. Measures are ordered according to their distance to degree

	k-core	Harm.	TC	Shapley(G1)	LTC(0.7)	PR	GDD(0.05)	Deg.
k-core	1000	42	47	31	80	63	66	127
Harm.		1000	224	397	603	468	348	422
TC			1000	281	295	373	460	453
Shapley (G1)				1000	516	723	477	456
LTC(0.7)					1000	676	485	601
PR						1000	664	701
GDD(0.05)							1000	761
Deg.								1000

Table 6 Average and median date of registration of the top 1000 nodes of various centrality measures on the social networks iWiW and Pokec. Registration date is measured in number of days from the kickoff of the network. Last row shows the average/median on the whole network

iWiW			Pokec		
Centrality	Average	Median	Centrality	Average	Median
TC	1554.402	1584	TC	2928.189	2891.5
Degree	1657.128	1645	GDD(0.05)	3057.238	3038
GDD(0.05)	1648.546	1646.5	k-core	3145.988	3160.5
LTC(0.7)	1673.296	1661	Degree	3169.621	3169
Harmonic	1670.558	1662.5	PageRank	3274.957	3320
PageRank	1695.087	1669	Harmonic	3300.377	3356
k-core	1753.594	1704	Shapley (G1)	3357.282	3386.5
Shapley (G1)	1759.33	1707	LTC (0.7)	3418.432	3435
Whole graph	1911.242	1813	Whole graph	3163.550	3192

Fig. 1 iWiW dates of registration of top 1000 users of various network centralities. Measures are ordered by the median day of registration



median, last row shows the average and the median for all nodes in the network.

In case of iWiW all measures performed well, that is, all averages/medians are below the network average/median. The Top Candidate (TC) method proved to be the best, with an average date of registration 7% lower than that of the next best centrality, Degree, and almost 20% lower than the network average.

TC retains its first place on Pokec as well, though with smaller margins. It performs 4.3% better than the next best, GDD, and 7.5% better than the network average. Note that,

the centralities showed much more volatility: Five out of the eight performed worse than the network average.

The results seem to be consistent. TC, GDD and Degree are among the first four, while Harmonic centrality, PageRank and Shapley (G1) lag behind on both networks. Only LTC and k-core showed varying results.

The average and median date of registration are, in themselves, imperfect indicators of performance. Due to their extreme risk-aversion, laggards would almost surely refuse to participate in a campaign, while individuals belonging to the early majority might be persuaded with, e.g., a small financial reward. Hence, we need to take a look at the whole distribution to evaluate the measures.

In case of iWiW, the field is mostly even (Fig. 1). TC is the only centrality that sticks out of the crowd, consistently outperforming the other measures in innovator and early adopter category, while also having the fewest laggards and late majority.

Although the performances are more nuanced in Pokec,

TC is still the best (Fig. 2). In case of innovators, its performance is on par with the other measures. This is perhaps due to the fact that very few individuals fell into this category. It has more early adopters and early majority and less late majority than any other centrality, while in laggards category, it is the second best. GDD also shows some very promising results.

Assuming that (i) A marketing message or a product sample will only incite innovators or early adopters, and that (ii) These two groups have their greatest influence on

Fig. 2 Pokec dates of registration of top 1000 users of various network centralities. Measures are ordered by the median day of registration

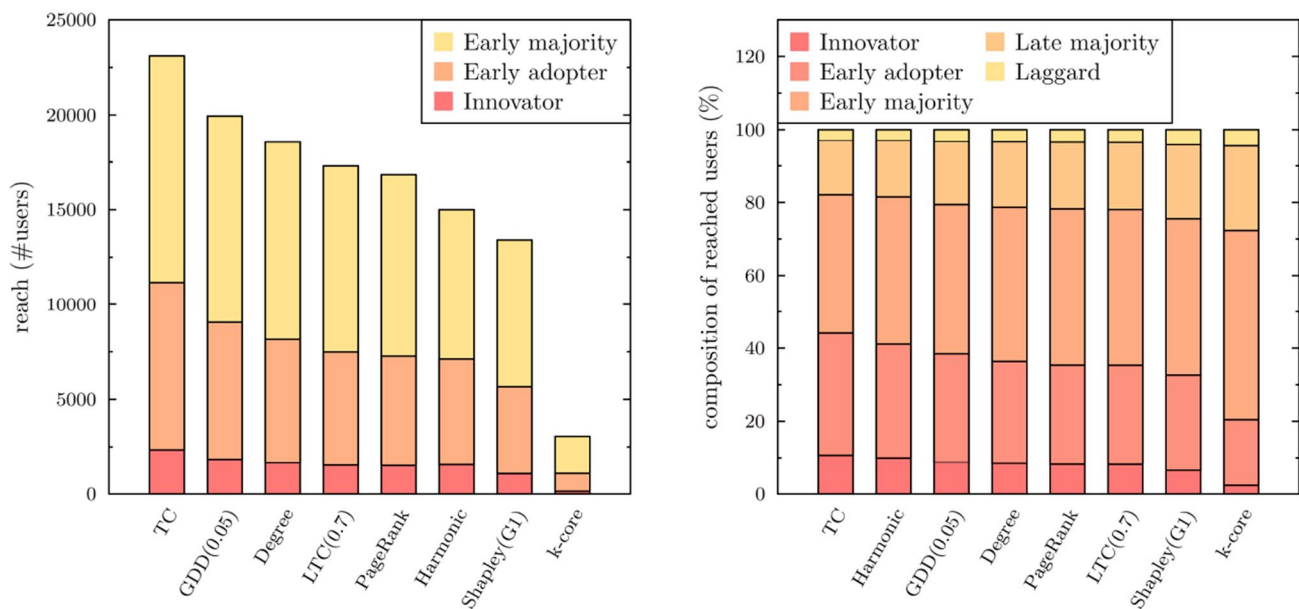
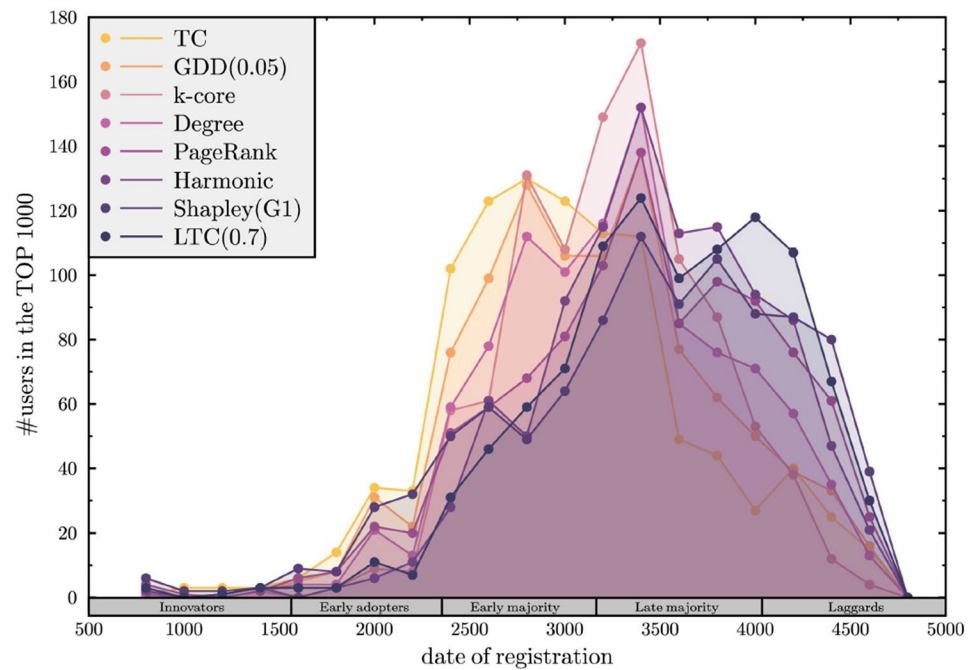


Fig. 3 Reach of innovators and early adopters among the top 1000 nodes of different centralities on iWiW

like-minded groups and on early majority, it is worth to restrict our attention to these two groups and their interactions with their neighbors. Figures 3 and 4 show the net reach of innovators and early adopters among the top 1000. The bar graph on the left depicts how many innovators, early adopters and early majority they reach not counting themselves. This illustrates the indirect impact of the campaign.

The bar graphs on the right hand side show the composition of their reach.

Note that, TC only comes out as a winner if these two assumptions hold—the bulk reach of, e.g., PageRank, that includes late majority and laggards as well, is much larger than that of TC. Thus, on a conventional linear threshold or independent cascade simulation, PageRank would outperform TC. However, by omitting these two assumptions,

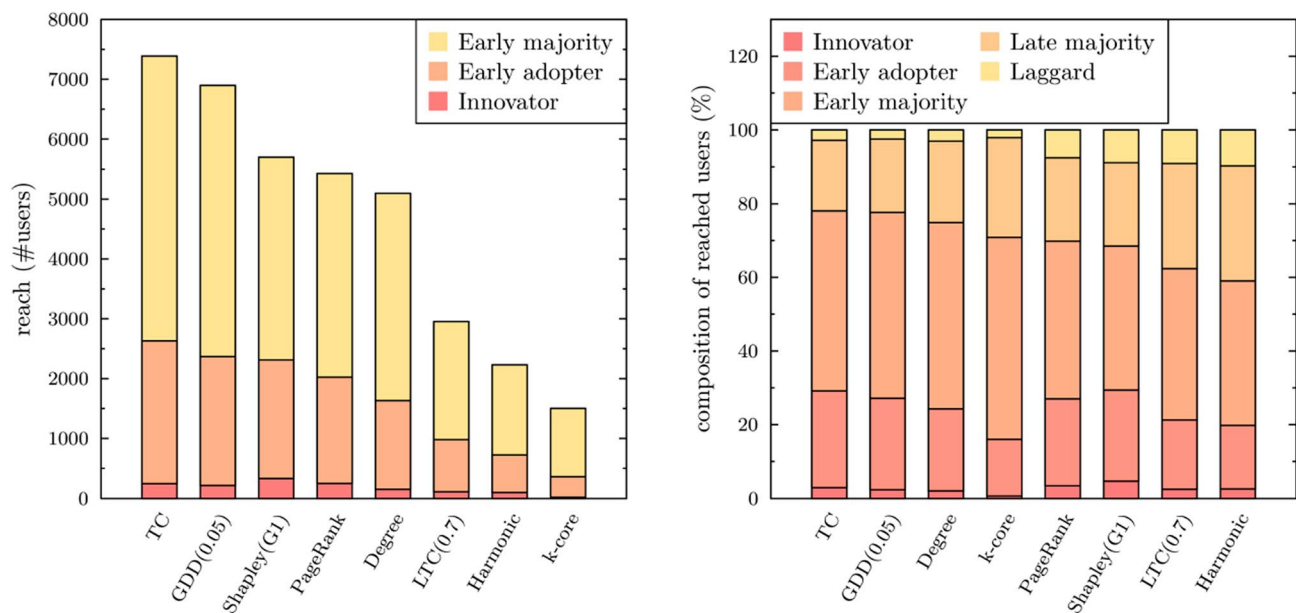


Fig. 4 Reach of innovators and early adopters among the top 1000 nodes of different centralities on Pokec

we oversimplify the diffusion model and assign inaccurate prediction power to the tested algorithms.

5 Conclusion

Innovators and early adopters are not abstract theoretical constructions, but groups that can be found on social networks as node clusters with distinct connection preferences. Consequently, they can be identified by observing the network structure. The top choices of some network centralities include more innovators and early adopters than others. Since these two groups play an essential role in innovation spreading, such network centralities might be more effective in real marketing campaigns.

Influence maximization aims to find the most influential nodes on the network. In the past two decades, myriads of clever heuristics were invented to optimize this computationally difficult task. Usually, these algorithms are validated via computer simulations with little care about what a real diffusion would look like. In real life, targeted agents often refuse to participate in the campaign. The underlying reasons are manifold, but most prominently agents differ in their risk attitudes. No matter how central a node is if it is risk-averse, unwilling to try the advertised product or commit to it openly.

Simulations also commonly ignore network homophily which can have serious impact on how a cascade unfolds. Both social networks presented here show strong patterns of homophily (Tables 1 and 2).

We tested eight different network centralities on two social networks where data about the date of registration were available. This allowed us to rank the centralities by their ability to identify innovators and early adopters. A novel expert selection algorithm, the Top Candidate method (TC), consistently outperformed every other method. To a smaller extent, Generalized Degree Discount and Degree were also effective.

A possible explanation of the success of the Top Candidate ranking is that individuals with high socioeconomic status and opinion leadership qualities—two traits that are associated with innovators and early adopters—are perceived as experts in society. Since the Top Candidate method is specifically designed to identify experts, it is a small wonder, that it finds more innovators and early adopters than other measures. The Top Candidate ranking is derived by the different parametrizations of the Top Candidate method. For a fixed parameter, the Top Candidate method outputs a list of individuals that form a stable set—the underlying idea is that experts are much more efficient in recognizing each other than amateurs, thus the selected individuals must support each other. This property resembles to assortativity and might be the reason why the method is successful in identifying such highly assortative sets as innovators and early adopters. Another possible explanation is that TC identifies more market mavens, who are also crucial in innovation spreading and widely acknowledged as experts.

The results may be interesting for practitioners of various fields. Computer scientists often test their heuristics with simulations on either the linear threshold or

the independent cascade models. In light of the results, the accuracy of these experiments can be improved by redesigning the threshold and propagation probability distributions. There are already a few papers that study how to obtain sensible propagation probabilities for the independent cascade model but less attention was given to node thresholds, and no papers take into account Roger's adopter classification when calibrating diffusion variables.

For marketing specialists, the practical lessons of this paper are that aiming for experts in a campaign might be a rewarding strategy, and that the Top Candidate method is an excellent tool for finding them.

6 Limitations and future research

In our study, we implicitly assumed that date of registration is a good proxy for innovativeness. Users that are keen to connect to the social network at an early stage can be reasonably categorized as innovators or early adopters at least regarding products and services related to social media. However, it is unclear how general the area is where this innovativeness applies. Opinion leaders tend to be monomorphic in nature, meaning they exercise their influence in a domain specific manner (Flynn et al. 1996; Doumit et al. 2011). The further we are from the initial product, the less certain we can be about their behavior. The same users might be innovative in information technology, but we cannot meaningfully say anything about their attitudes toward unrelated subjects like food and fashion. Thus, researchers of innovation diffusion should always look for additional characteristics beside network position.

The high level of assortativity that is observable in social networks calls for the revision of diffusion models. It would be expedient to test variants of the linear threshold and independent cascade models that account for homophily.

The good performance of Top Candidate method suggests that at least some fraction of the innovators and early adopters are considered experts in society. However, the rich characterization of innovators and early adopters does not traditionally include the 'expert' label. An interesting sociometric line of research would be to explore the relationship between early adoption (or in general innovativeness) and perceived expertise.

Acknowledgements The authors acknowledge financial help received from National Research, Development and Innovation Office, grant numbers K 138945 (Sziklai) and KH 130502 (Lengyel). Balázs R. Sziklai is the grantee of the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, supported by the ÚNKP-22-5 New National Excellence Program of the Ministry for Culture and

Innovation from the source of the National Research, Development and Innovation Fund.

Funding Open access funding provided by Corvinus University of Budapest.

Declarations

Conflict of interest The authors have no conflicts of interest to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Anwar MS, Saveski M, Roy D (2021) Balanced influence maximization in the presence of homophily. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining, WSDM '21, Association for Computing Machinery, New York, NY, USA, pp 175–183
- Aral S, Dhillon PS (2018) Social influence maximization under empirical influence models. *Nat Hum Behav* 2(6):375–382
- Aral S, Walker D (2012) Identifying influential and susceptible members of social networks. *Science* 337(6092):337–341. <https://doi.org/10.1126/science.1215842>
- Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO (2013) The diffusion of microfinance. *Science* 341(6144):1236498
- Bass FM (1969) A new product growth for model consumer durables. *Manage Sci* 15(5):215–227
- Bokányi E, Novák M, Jakobi Á, Lengyel B (2022) Urban hierarchy and spatial diffusion over the innovation life cycle. *Royal Soci Open Sci* 9(5):211038
- Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. *J Math Sociol* 2(1):113–120
- Bóta A, Csernenszky A, Györfy L, Kovács G, Krész M, Pluhár A (2015) Applications of the inverse infection problem on bank transaction networks. *CEJOR* 23(2):345–356. <https://doi.org/10.1007/s10100-014-0375-2>
- Bóta A, Krész M, Pluhár A (2016) Estimation of edge infection probabilities in the inverse infection problem. Springer International Publishing, Cham, pp 17–36
- Brancheau JC, Wetherbe JC (1990) The adoption of spreadsheet software: testing innovation diffusion theory in the context of end-user computing. *Inf Syst Res* 1(2):115–143
- Calió A, Tagarelli A (2021) Attribute based diversification of seeds for targeted influence maximization. *Inf Sci* 546:1273–1305. <https://doi.org/10.1016/j.ins.2020.08.093>
- Centola D, Macy M (2007) Complex contagions and the weakness of long ties. *Am J Sociol* 113(3):702–734
- Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining,

- KDD '09, Association for Computing Machinery, New York, NY, USA, pp 199–208
- Cho Y, Hwang J, Lee D (2012) Identification of effective opinion leaders in the diffusion of technological innovation: a social network approach. *Technol Forecast Soc Chang* 79(1):97–106. <https://doi.org/10.1016/j.techfore.2011.06.003>
- Csató L (2017) Measuring centrality by a generalization of degree. *CEJOR* 25(4):771–790
- De Arruda GF, Barbieri AL, Rodriguez PM, Rodrigues FA, Moreno Y, da Fontoura Costa L (2014) Role of centrality for the identification of influential spreaders in complex networks. *Phys Rev E* 90(3):032812
- Dedehayir O, Ortt RJ, Riverola C, Miralles F (2017) Innovators and early adopters in the diffusion of innovations: a literature review. *Int J Innov Manag* 21(08):1740010
- Doumit G, Wright FC, Graham ID, Smith A, Grimshaw J (2011) Opinion leaders and changes over time: a survey. *Implement Sci* 6(1):117. <https://doi.org/10.1186/1748-5908-6-117>
- Eastlick M, Lotz S (1999) Profiling potential adopters and non-adopters of an interactive electronic shopping medium. *Int J Retail Distrib Manag* 27(6):209–223
- Feick LF, Price LL (1987) The market maven: a diffuser of marketplace information. *J Mark* 51(1):83–97
- Flynn LR, Goldsmith RE, Eastman JK (1996) Opinion leaders and opinion seekers: two new measurement scales. *J Acad Mark Sci* 24(2):137. <https://doi.org/10.1177/0092070396242004>
- Goldsmith RE, Flynn LR, Goldsmith EB (2003) Innovative consumers and market mavens. *J Market Theory Pract* 11(4):54–65. <https://doi.org/10.1080/10696679.2003.11658508>
- Goyal A, Bonchi F, Lakshmanan LV (2010) Learning influence probabilities in social networks. In: *Proceedings of the 3rd ACM international conference on web search and data mining, WSDM '10*, Association for Computing Machinery, New York, NY, USA, pp 241–250
- Granovetter M (1978) Threshold models of collective behavior. *Am J Sociol* 83:489–515
- Hajdu L, Bóta A, Krész M, Khani A, Gardner LM (2020) Discovering the hidden community structure of public transportation networks. *Netw Spat Econ* 20(1):209–231. <https://doi.org/10.1007/s11067-019-09476-3>
- Jung K, Heo W, Chen W (2012) Irie: Scalable and robust influence maximization in social networks. In: *2012 IEEE 12th international conference on data mining*, pp 918–923
- Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. In: *Proceedings of the 9th ACM SIGKDD international conference on knowledge discovery and data mining, KDD '03*, Association for Computing Machinery, New York, NY, USA, pp 137–146
- Kitsak M, Gallos LK, Havlin S, Liljeros F, Muchnik L, Stanley HE, Makse HA (2010) Identification of influential spreaders in complex networks. *Nat Phys* 6:888–893
- Lengyel B, Bokányi E, Di Clemente R, Kertész J, González MC (2020) The role of geography in the complex diffusion of innovations. *Sci Rep* 10(1):1–11
- Lynn T, Muzellec L, Caemmerer B, Turley D (2011) Social network sites: early adopters' personality and influence. *J Prod Brand Manag* 26:42–51
- Marchiori M, Latora V (2000) Harmony in the small-world. *Physica A* 285(3):539–546
- McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Ann Rev Sociol* 27(1):415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- Michalak TP, Aadithya KV, Szczepański PL, Ravindran B, Jennings NR (2013) Efficient computation of the shapley value for game-theoretic network centrality. *J Artif Intell Res* 46:607–650
- Moore GA (2014) *Crossing the Chasm: marketing and selling disruptive products to mainstream customers*, 3rd edn. HarperCollins Publishers, New York
- Muller E, Yogeve G (2006) When does the majority become a majority? empirical analysis of the time at which main market adopters purchase the bulk of our sales. *Technol Forecast Soc Chang* 73(9):1107–1120. <https://doi.org/10.1016/j.techfore.2005.12.009>
- Newman ME (2002) Assortative mixing in networks. *Phys Rev Lett* 89(20):208701
- Page L, Brin S, Motwani R, Winograd T (1999) The pagerank citation ranking: bringing order to the web. Technical Report 1999-66, Stanford InfoLab. Previous number = SIDL-WP-1999-0120
- Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A (2015) Epidemic processes in complex networks. *Rev Mod Phys* 87(3):925
- Plötz P, Schneider U, Globisch J, Dütschke E (2014) Who will buy electric vehicles? identifying early adopters in Germany. *Transp Res Part A: Policy Pract* 67:96–109. <https://doi.org/10.1016/j.tra.2014.06.006>
- Qiang Z, Pasiliao EL, Zheng QP (2019) Model-based learning of information diffusion in social media networks. *Appl Netw Sci* 4(1):1–6
- Rahimkhani K, Aleahmad A, Rahgozar M, Moeini A (2015) A fast algorithm for finding most influential people based on the linear threshold model. *Expert Syst Appl* 42(3):1353–1361. <https://doi.org/10.1016/j.eswa.2014.09.037>
- Riquelme F, Gonzalez-Cantergiani P, Molinero X, Serna M (2018) Centrality measure in social networks based on linear threshold model. *Knowl-Based Syst* 140:92–102
- Rogers EM (1962) *Diffusion of innovations*, 1st edn. Free Press of Glencoe, New York
- Rogers EM (2003) *Diffusion of innovations*, 5th edn. Free Press, New York, NY
- Seebauer S (2015) Why early adopters engage in interpersonal diffusion of technological innovations: an empirical study on electric bicycles and electric scooters. *Transp Res Part A: Policy Pract* 78:146–160
- Seidman SB (1983) Network structure and minimum degree. *Social Networks* 5(3):269–287
- Selfhout M, Burk W, Branje S, Denissen J, van Aken M, Meeus W (2010) Emerging late adolescent friendship networks and big five personality traits: a social network approach. *J Pers* 78(2):509–538
- Shapley LS (1953) A value for n-person games. *Ann Math Stud*. <https://doi.org/10.1515/9781400881970-018>
- Suri NR, Narahari Y (2008) Determining the top-k nodes in social networks using the Shapley value. In: *Proceedings of the 7th international joint conference on autonomous agents and multiagent systems, Vol. 3, AAMAS '08*, Richland, SC, pp 1509–1512
- Sziklai BR (2018) How to identify experts in a community? *Int J Game Theory* 47:155–173
- Sziklai BR (2021) Ranking institutions within a discipline: the steep mountain of academic excellence. *J Informet* 15(2):101133
- Toole JL, Cha M, González MC (2012) Modeling the adoption of innovations in the presence of geographic and media influences. *PLoS ONE* 7(1):e29528
- Török J, Kertész J (2017) Cascading collapse of online social networks. *Sci Rep* 7(1):1–8
- Valente TW (1996) Social network thresholds in the diffusion of innovations. *Social networks* 18(1):69–89
- Wang X, Lan Y, Xiao J (2019) Anomalous structure and dynamics in news diffusion among heterogeneous individuals. *Nat Hum Behav* 3(7):709–718. <https://doi.org/10.1038/s41562-019-0605-7>
- Wang X, Zhang X, Zhao C, Yi D (2016) Maximizing the spread of influence via generalized degree discount. *PLoS ONE* 11(10):1–16

- Watts DJ (2002) A simple model of global cascades on random networks. *Proc Natl Acad Sci* 99(9):5766–5771
- Weng L, Menczer F, Ahn YY (2014) Predicting successful memes using network and community structure
- Was T, Skibski O (2018) Axiomatization of the pagerank centrality. In: *Proceedings of the 27th international joint conference on artificial intelligence, IJCAI-18*, pp 3898–3904
- Young HP (1985) Monotonic solutions of cooperative games. *Internat J Game Theory* 14(2):65–72. <https://doi.org/10.1007/BF01769885>
- Zijlstra T, Durand A, Hoogendoorn-Lanser S, Harms L (2020) Early adopters of mobility-as-a-service in the Netherlands. *Transp Policy* 97:197–209. <https://doi.org/10.1016/j.tranpol.2020.07.019>
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.