Opinion Leaders in Star-Like Social Networks: A simple Case?

Michael Spranger, Florian Heinke, Hanna Siewerts, Joshua Hampl and Dirk Labudde
University of Applied Sciences Mittweida
Mittweida, Germany
Email: {*name.surname*}@hs-mittweida.de

***Abstract*—In recent years, the automated, efficient and sensitive monitoring of social networks has become increasingly important for criminal investigations and crime prevention. Previously, we have shown that the detection of opinion leaders is of great interest in forensic applications. In the present study, it is argued that state of the art opinion leader detection methods have weaknesses if networks exhibit star-like social graph topology, whereas these topologies result from the interaction of users with similar interests. This is typically the case for Facebook pages of political organizations. In these cases, the underlying topologies are highly focused on one (or only a few) central actor(s) and lead to less meaningful results by classic measures of node centrality commonly used for leader detection. The presents study examines these aspects closer and exemplifies them with the help of data collected from the Facebook page of a German political party for five consecutive months. Furthermore, a quantitative indicator for describing star-like network topologies is introduced and discussed. This measure can be of great value in assessing the applicability of established leader detection methods. Finally, a modified LeaderRank score is proposed – the CompetenceRank – which aims to address discussed problems.**

***Keywords*–Forensic; Opinion Leader; Graph Theory.**

I. Introduction

The detection of opinion leader has been discussed extensively in the past few years. Based on the work by Katz [7] many approaches have been presented. In this paper it will be shown that in some situations these approaches do not capture the core of the problem and as a result lead to an inaccurate assessment of opinion leadership. This section shall give a brief introduction to the field in which such situations occur as well as an overview of topology-based approaches and finishes with the scope and structure of the paper.

A. General Motivation

Analyzing social networks has become an important tool for investigators, intelligence services and decision makers of police services. The information gained this way can be used to solve crimes by searching for digital evidence that relates to the crime in the real world. Additionally, methods of predictive policing can help to organize police missions as was shown in [1]–[3]. The detection of opinion leaders in social networks is an important task for different reasons. On the one hand, owners of influential profiles are often also influential in the offline world. Knowing these people helps to determine the direction of an investigation or more concretely to target persons of interest. On the other hand, as was suggested in previous work [3], it might be of interest to contact these profiles by means of chatbots to gain access into closed groups in an effort to gather important information for intelligence services. Intuitively, opinion leaders, when considered as nodes with high structural importance, can be detected with the help of centrality measures. However, different kinds of influence in a network have to be distinguished. Nodes can have a great influence as corresponding actors are able to spread information fast and widely in a network, or they can have a great influence because they write something of importance, which attracts many other users in the network to respond.

B. Leader Detection by means of Network Centrality Measures

In the literature, one can mainly find centrality measures for the former type of influence. For example, highly active profiles can be recognized using degree centrality, meaning, the relative number of outgoing edges of a node. These profiles are represented by nodes with a high degree centrality which are especially useful to spread information in a network due to their high interconnectedness.

In this context, the closeness centrality – the inverse of the mean of the shortest path of a node to any other node in the network – is even more effective. It describes the efficiency of the dissemination of information of a certain node.

Furthermore, the betweenness centrality of a certain node, which is defined as the number of shortest paths between two nodes that cross this node, describes the importance of this node for the dissemination of information in a network. Therefore, the higher the betweenness centrality of a node, the greater its importance for the exchange of information in a network.

Moreover, the eigenvector centrality of a node is defined as the principle eigenvector of the adjacency matrix of a network. In contrast to the measures discussed beforehand, PageRank [4], as one of the best measures of node centrality, does not only consider the centrality of the node itself, yet also of its neighboring nodes.

As part of the opinion leader detection research, LeaderRank [5] was introduced as a further development of PageRank in order to find nodes that spread information further and faster. However, all of these centrality measures consider nodes that are involved in the dissemination of information mainly based on their activity. For the purpose of the intended usage, users who achieve high impact through what they have written are of much greater interest. Thus, similar to the citation of papers and books and its impact on the author's reputation, the importance of a node has to be higher when it reaches a high number of references and citations with low activity.

Interestingly, Li et al. considered the so-called node spreadability as the ground truth for quantifying node importance in a subsequent study [6]. Node spreadability is based on a straightforward Susceptible-Infected-Removed (SIR) infection

 ISBN: 978-1-61208-654-5

model from which the expected number of infected nodes upon initially infecting the node in question is estimated. However, this expected number can only be estimated from simulation, which furthermore is dependent on the parameterization of the SIR model. In this respect, all centrality measures can be considered as heuristic approximations of node spreadability.

C. Scope and Structure of the Paper

In this case study, we discuss problems that can arise when aiming to detect opinion leaders in social networks yielding highly central topologies similar to star graphs. Examples for such networks are group pages on Facebook or vk.com where user interactions and activities are mostly triggered by and focused on posts made by the page owner. In such cases, the page owner – a trivial leader in the sense of centrality measures discussed above – acts as a score aggregator and can thus lead to distorted scoring, which can eventually be adverse in the context of opinion leader detection. In this case, classic centrality measures can be considered inappropriate. Based on interactions of users of the political Facebook page “DIE LINKE” tracked for five consecutive months (January - May 2017), this problem is illustrated. We further introduce the LeaderRank skewness as a quantitative measure of aggregator-induced distorted LeaderRank scoring, which in experiments show to be superior to network entropy with respect to expressiveness. Finally, a modified LeaderRank score, we refer to as CompetenceRank, is introduced which is proposed to be suitable for opinion leader detection in such networks.

The paper is structured as follows: in Section II, a brief literature overview on the topic of opinion leader detection is given, followed by a summary of the LeaderRank algorithm. Issues of LeaderRank scoring in star-shaped network topologies are discussed in Section III, including the deduction and definition of LeaderRank skewness. In Section IV, the social network dataset in question is discussed. The CompetenceRank is introduced in Section V. We finally give a conclusion as well as an overview on future work in Section VI.

II. Detection of Opinion Leaders

Opinion leaders in the context of the intended analysis of social networks are individuals, who exert a significant amount of influence on the opinion and sentiment of other users of the network through their actions or by what they are communicating. In social sciences the term ’opinion leader’ was introduced before 1957 by Katz and Lazarsfeld’s research on diffusion theory [7]. Their proposed two-step flow model retains validity in the digital age, especially in the context of social media.

Katz et al. assume that information disseminated in the Social Network is received, strengthened and enriched by opinion leaders in their social environment. Each individual is influenced in his opinion by a variety of heterogeneous opinion leaders. This signifies, that the opinion of an individual is mostly formed by its social environment. In 1962, Rogers referenced these ideas and defined opinion leader as follows:

> “Opinion leadership is the degree to which an individual is able to influence informally other individuals’ attitudes or overt behavior in a desired way with relative frequency.” [8, p. 331]

For the present study, one important question to answer is what influence means, or rather how to identify an opinion leader or how the influencer can be distinguished from those being influenced. Katz defined the following features [7]:

1) personification of certain values,
2) competence,
3) strategic social location.

One approach to identify opinion-leaders is to extract and analyze the content of nodes and edges of networks to mine leadership features. For instance, the sentiment of communication pieces can be analyzed to detect the influence of their authors, as shown by Huang et. al., who aim to detect the most influential comments in a network this way [9]. Another strategy is to perform topic mining to categorize content and detect opinion leaders for each topic individually, as opinion leadership is context-dependent [7] [10]. For this purpose, Latent Dirichlet Allocation (LDA) [11] can be used, as seen in the work of [12].

In this study, we considered the implementation of content-based methods problematic, as texts in social networks mostly lack correct spelling and formal structure, which impairs such methods’ performance. Additionally, leaders can be identified by analyzing the flow of information in a network. By monitoring how the interaction of actors evolves over time, one can identify patterns and individuals of significance within them. To achieve this, some model of information propagation is required, such as Markov processes employed by [13] and the probabilistic models proposed by [14]. These interaction-based methods consider both topological features and their dynamics over time.

We utilized methods, which are solely based on a network’s topology, therefore, consider features, such as node degree, neighborhood distances and clusters, to identify opinion leaders. One implementation of this is the calculation of node centrality. The underlying assumption is that the more influence an individual gains, the more central it is in the network. Which centrality measure is most suitable is dependent on the application domain. We judged eigenvector centrality to be most adequate. One of the most popular algorithms is Google’s PageRank algorithm [4]. The application of PageRank for the purposes of opinion leader detection has seen merely moderate success [15] [16]. With LeaderRank scores, Lü et al. advocate further development and optimization of this algorithm for social networks, and have achieved surprisingly good results [5]. Herein, users are considered as nodes and directed edges as relationships between opinion leaders and users. All users are also bidirectionally connected to a ground node, which ensures connectivity as well as score convergence. In short, the algorithm is an iterative multiplication of a vector comprised by per-node scores $s_i(t)$ at iteration step t with a weighted adjacency matrix until convergence is achieved according to some convergence criteria. Initially, at iteration step t_0, all vertex scores are set to $s(0) = 1$, except for the ground node score which is initialized as $s_g(0) = 0$. Equation (1) describes LeaderRank algorithm as a model of probability flow through the network, where $s_i(t)$ indicates the score of a node i at iteration step t.

$$s_i(t+1) = \sum_{j=1}^{N+1} \frac{a_{ji}}{k_j^{out}} s_j(t) \tag{1}$$

Depending on whether or not there exists a directed edge from node i to node j, the value 1 respectively 0 is assigned to a_{ij}. k_i^{out} describes the number of outgoing edges of a node. The final score is obtained as the score of the respective node at the convergence step t_c and the obtained ground node score, as shown in (2). At t_c, equilibration of LeaderRank scores towards a steady state can be observed.

$$S_i = s_i(t_c) + \frac{s_g(t_c)}{N} \tag{2}$$

The advantage of this algorithm compared to PageRank is that the convergence is faster and above all that nodes, that spread information faster and further, can be found. In later work, for example, by introducing a weighting factor, as in [6] or [17], susceptibility to noisy data has been further reduced and the ability to find influential distributors (hubs) of information has been added.

III. Issues with LeaderRank

The LeaderRank algorithm can be understood as a reversion of a discrete model of diffusion. In that sense, the initialization $s_i(0) = 1$ at t_0 can be interpreted as assigning a uniform concentration distribution of some virtual compound which in the processes is re-distributed according to the model. In that respect, central actors showing the highest activity in star-like networks can induce score aggregation and migration towards their central nodes as well as their adjacent nodes, whereas nodes in the 'peripheral region' of the network become inadequately represented by their scores. One can thus hypothesize that ranked lists obtained from LeaderRank scores can not be considered meaningful if a given network in question exhibits star-topological topology.

Furthermore, the LeaderRank emphasizes the strategic social location of a user, whereas their competence is not considered. In star-shaped network topologies, high centralities of only a fraction of nodes leads to a heavily skewed LeaderRank score distribution.

In this case study, the network under investigation shows an even more extreme case of star topology in which the owner of the political Facebook page 'DIE LINKE' acts solely as the central actor (for more information see Section IV). In contrast, one could argue that someone is more important if any activity generates a high number of responses. Such a case is regularly given by political networks, which are dominated by the central node of the page owner. Thus, a straightforward modification of the LeaderRank score is proposed in Section V which addresses the imbalance the LeaderRank algorithm yields in such networks.

In the following paragraph a quantitative measure of LeaderRank distribution skewness is proposed, which could aid to ensure proper applicability of the LeaderRank algorithm for any given network. This measure is further compared to the classic measure of network entropy. Tests on simulated data show the LeaderRank skewness to be superior to network entropy with respect to topological changes.

A. Definition of LeaderRank Distribution Skewness

Let $LR = \{lr_1, ..., lr_i, ..., lr_N\}$ be the LeaderRank values of all nodes. Further, $\overline{lr}$ and sd_{LR} denote the arithmetic mean and standard deviation of LR. Based on the z-scaled LeaderRank values (3), the skewness ν of the LeaderRank distribution is calculated as shown in (4).

$$z(lr_i) = \frac{lr_i - \overline{lr}}{sd_{LR}} \tag{3}$$

$$\nu_{LR} = \left| \frac{1}{N} \sum_i z(lr_i)^3 \right| \tag{4}$$

As discussed above, score distribution skewness is correlated with network topology. Yet, normalization of computed skewness is required in order to make predications about the topology and whether a star-like topology is present. Thus upper and lower bounds, ν_{min} and ν_{max}, are needed. In this paragraph, derivation of both bounds are given.

Trivially, ν converges to the lower bound – the theoretical minimum ($\nu = 0$) – in almost regular graphs. Such graphs are regular graphs with one edge being removed. With N being sufficiently large, the supposition that $lr_i \approx lr_j$ of any pair of randomly selected vertices v_i and v_j holds true and a limit of $\lim_{sd_{LR} \to 0} \nu = 0$ can be assumed. In regular graphs however all LeaderRank scores are equal by definition, resulting to $sd_{LR} = 0$ and ν being undefined in this case.

In contrast, ν is equal to the theoretical maximum if the network graph exhibits a strictly star-shaped topology. Let lr_c be the value of the LeaderRank of the central vertex of such a network. The LeaderRank values of any randomly selected pair of vertices v_i and v_j with $v_i, v_j \neq v_c$ are then not distinguishable, i. e., $lr_i = lr_j$, according to the LeaderRank's definition. Furthermore, the total LeaderRank generally sums up to N, $LR_{tot} = \sum_{i=1}^{N} lr_i = N$ (which leads to $\overline{lr} = 1$ for every graph). Given the central node lr_c, each lr_i can thus be calculated as shown in (5).

$$lr_i = \frac{N - lr_c}{N - 1} \tag{5}$$

If lr_c is known, the set of LeaderRank values $\{lr_c, lr_2, ...lr_i, ...lr_N\}$ and the resulting ν_{max} can be derived. For star graphs of any size N, a linear correlation exists between lr_c and N ($lr_c \approx 0.20N + 0.66, R = 1.0$, data not shown). The upper skewness bound ν_{max} can thus be readily computed. Subsequently, for any irregular network graph LeaderRank skewness can be calculated and normalized subsequently using a min-max normalization as denoted in (6), whereas ν_{min} can be assumed as 0 as discussed above.

$$\hat{\nu} = \frac{\nu - \nu_{min}}{\nu_{max} - \nu_{min}} = \frac{\nu}{\nu_{max}} \tag{6}$$

B. Detection of star topology

LeaderRank skewness $\hat{\nu}$ can be utilized to indicate adverse leader ranking by means of LeaderRank scores. In this section, we compare ν to the classic measure of network entropy (denoted as H in the following text). In order to allow direct comparison to $\hat{\nu}$ as well as to entropies computed from other graphs, H is required to be normalized analogously to $\hat{\nu}$. In this subsection, we give a brief overview on how normalization can be conducted.

Let A be the adjacency matrix of a network with N vertices, where each element $a_{ij} := 1$, if there exists a directed edge e_{ij} between adjacent vertices v_i and v_j. Each element of

 ISBN: 978-1-61208-654-5

the principal diagonal a_{ii} is defined as $a_{ii} := \deg(v_i)$ and thus corresponds to the degree – the sum of the incoming and outgoing links – of vertex v_i. The trace of A is defined as the sum of all elements of the principal diagonal: $tr(A) = \sum_{i=1}^{N} a_{ii}$. The formalism for graph entropy used by Passerini and Severini $S(\rho) = -tr(\rho \log_2 \rho)$ [18] is based on the von Neumann entropy and can be adapted as follows:

$$\begin{aligned} S(\rho) &= -tr(\rho \log_2 \rho) \\ &= -\sum_{i=1}^{N} \rho_i \log_2 \rho_i \\ &= -\sum_{i=1}^{N} \frac{a_{ii}}{tr(A)} \log_2 \frac{a_{ii}}{tr(A)} \\ &= -\sum_{i=1}^{N} \frac{\deg(v_i)}{\sum\limits_{j=1}^{N} \deg(v_j)} \log_2 \frac{\deg(v_i)}{\sum\limits_{j=1}^{N} \deg(v_j)}. \end{aligned} \tag{7}$$

The matrix entropy describes the distribution of incoming and outgoing links in a graph. In a randomly generated graph one expects $\deg(v_i) \approx \deg(v_j)$. In this case the matrix entropy H approaches its maximum H_{max}. Graph entropy is thus only at a maximum if G is a regular graph where $\deg(v_i) = \deg(v_j) = D$. Because $\rho_i = D/DN = 1/N$ in a regular graph, one has H as shown in (8).

$$H = H_{max} = -\sum \rho_i \log_2 \rho_i = \log_2 N \tag{8}$$

In contrast, the minimum matrix entropy H_{min} is observable in networks showing star topology. The trace $tr(A)$ of such a graph corresponds to $2N-2$ and the degree of its central vertex is $\deg(v_c) = N-1$. Consequently, the entropy of the central vertex H_c is calculated as shown in (9).

$$H_c = -\frac{N-1}{2N-2} \log_2 \frac{N-1}{2N-2} = -\frac{1}{2} \log_2 \frac{1}{2} = 0.5. \tag{9}$$

The degree of any other vertex is $\deg(v_i) = 1$. Hence, the entropy of a graph constituted as a star is calculated as follows:

$$\begin{aligned} H &= H_{min} \\ &= 0.5 + \sum_{V \setminus v_c} -\frac{1}{2N-2} \log_2 \frac{1}{2N-2} \\ &= 0.5 + \frac{1}{2} \log_2(2N-2) \\ &= 1 + \frac{1}{2} \log_2(N-1). \end{aligned} \tag{10}$$

Normalized network entropy can be finally computed according to (11):

$$\hat{H} = \frac{H - H_{min}}{H_{max} - H_{min}}, \hat{H} \in [0,1] \tag{11}$$

In order to illustrate expressiveness of $\hat{H}$ and $\hat{\nu}$ with respect to the underlying network topology, a straightforward experiment was carried out in which synthetic networks exhibiting star topologies were continuously mutated over time, resulting in almost regular graphs after numerous generations. This simulated process thus yields a continuous change of the network topology for each graph. $\hat{H}$ and $\hat{\nu}$ were accordingly computed for every generation and tracked. The time series of both measures are shown in Figure 1. More precisely, simulations of topological change were conducted by starting with star graphs of fixed sizes ($N = 16, 32, 64, 128, 256$ and 512 vertices). In every generation, edges between every pair of vertices were randomly added and respectively removed. For each graph size, six runs were conducted in an effort to estimate variance.

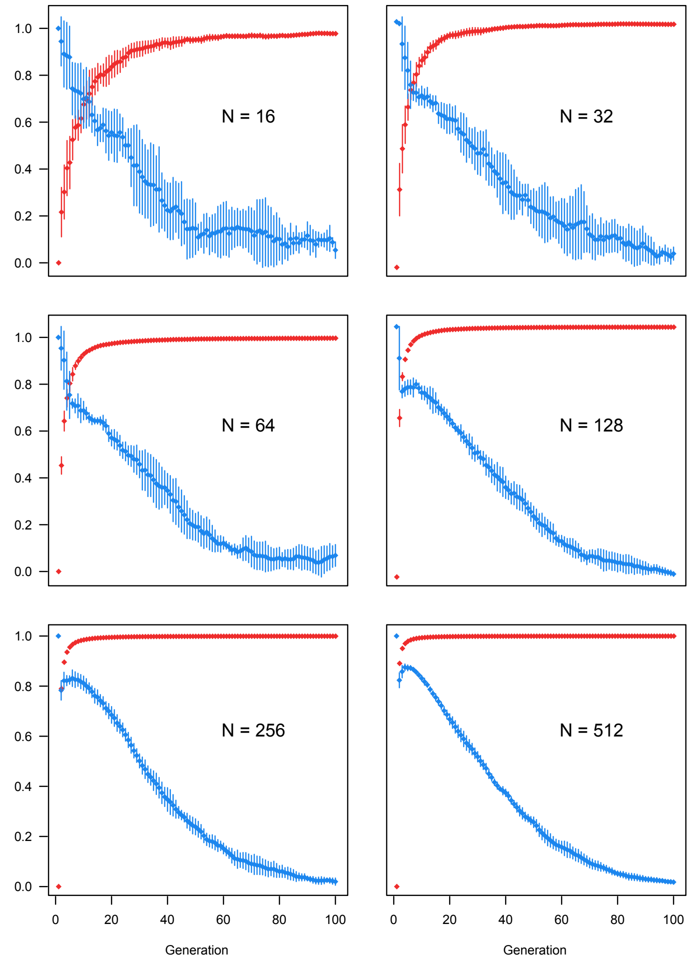


Figure 1. Simulation results of networks with various sizes N, whereas the red line represents $\hat{H}$, the blue line $\hat{\nu}$ and vertical bars indicate standard deviations.

As shown in Figure 1, both measures converged after 100 generations. All entropy trajectories show fast convergence compared to $\hat{\nu}$ trajectories, with the convergence time decreasing with increasing N. Although $\hat{\nu}$ yield larger variances (especially for $N \leq 32$), its slower convergence and qualitatively similar trajectories for all graph sizes N illustrates greater sensitivity to topological changes. In that respect matrix entropy loses significance with increasing graph size.

IV. DATASET

For this study, the structure of the Facebook page of the German party “DIE LINKE” was analyzed because it is a typical star-like topology with the page owner as a central node. This central node often has the highest activity, meaning the most in- and out-links. The communication on the page was explored over a period of five months, from January 2017 up until May 2017, whereas all posts, comments and replies were taken into account (see Table I).

TABLE I. SUMMARY OF THE DATA INCLUDING NORMALIZED ENTROPY AND SKEWNESS OF THE CONSIDERED NETWORKS.

month	actors	posts	comments	replies	$\hat{H}$	$\hat{\nu}_{LR}$
January	2,878	26	2,955	3,471	0.19	0.98
February	2,146	33	2,196	2,062	0.24	0.98
March	3,196	40	3,501	3,245	0.17	0.97
April	2,432	26	2,558	3,295	0.22	0.98
May	4,765	31	4,130	5,674	0.10	0.98
Epinions	75,879	n/a	n/a	n/a	0.65	0.07

During initial analysis of the dataset, it was observed that $12,031$ individuals were active during the five months. However, as shown in Figure 2, only 104 of these individuals were active in every single month. In general, it can be stated that users showed rather sparse and sporadic activity, with only a minority being recurrent users.

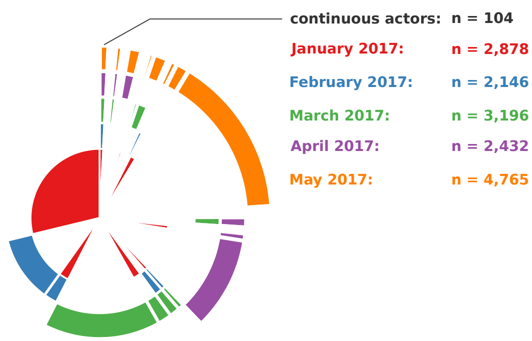


Figure 2. Sunburst chart of actor activity consisting of one radial segment for each user, whereas a user's segment in a time layer is left out if said user was observed to be inactive in that time period.

Figure 3 shows a comparison of two different network topologies. Each network represents the interaction of the users, in particular their communication, in a social network. The labels of the nodes of the users were anonymized using enumeration except of the central node in Figure 3a. This figure depicts the network of the Facebook page "DIE LINKE" from January 2017 as a graph in which the size of each node corresponds to the out-degree (number of out-links). As can be seen, the network is dominated by the central node of the page owner and thus is close to a star-shaped topology. In contrast, Figure 3b shows a part of the *Epinions* social network [19]. Due to the size of the network, it was necessary to limit the depiction by applying k-core ≥ 80, showing only the most active nodes. This network tends to be more decentralized, in other words, there is no node which dominates all others in terms of its degree.

Table I shows the normalized entropy and LeaderRank skewness of the "DIE LINKE" network, separately calculated for each month. It can be clearly seen, that obtained $\hat{H}$ values fluctuate over time, whereas LeaderRank skewness $\hat{\nu}_{LR}$ remains stable. For comparison, the *Epinions* social network [19] shows a considerably less skewed LeaderRank distribution, whereas normalized network entropy $\hat{H}$ is thus less expressive, as theoretically discussed in Section III.

V. CompetenceRank

To address the issues discussed in Section III, we present a competence-adjusted variant of the LeaderRank which down ranks nodes with a high amount of out-links in comparison to their in-links. The competence-adjusted LeaderRank, referred to as CompetenceRank, of a particular topic-specific opinion leader $CR(L_i)$ can be calculated as shown in (12).

$$CR(L_i) = \frac{LR(L_i)}{1 + \frac{k_i^{out}}{k_{total}^{out}} \cdot LR_{total}} \tag{12}$$

The CompetenceRank of a certain opinion-leader is calculated by dividing its original LeaderRank score LR_i by a fraction of cumulative sum of LeaderRank scores (which is equal to the number of users) defined by the node's share of network activity, and k^{out} being the number of outgoing links. By definition, LR_{total} – the sum of LeaderRank scores of all nodes in the social network graph – is equal to the number of nodes N. When considering regular graphs, one observes LeaderRank distribution skewness $\hat{\nu} = 0$ as well as $k_i^{out} = k_j^{out} = D$ for any pair of randomly chosen nodes v_i and v_j. Thus, $k_{total}^{out} = ND$. From this, the expression above can be conveniently rewritten as

$$CR(L_i) = \frac{LR(L_i)}{1 + \frac{D}{ND} \cdot N} = \frac{1}{2} LR(L_i). \tag{13}$$

We finally define the CompetenceRank based on the assumption that $LR(L_i) = CR_i$ in regular graphs, which is thus simply achieved by multiplying the expression in (13) by 2. Henceforth,

$$CR(L_i) = 2 \cdot \frac{LR(L_i)}{1 + \frac{k_i^{out}}{k_{total}^{out}} \cdot N}. \tag{14}$$

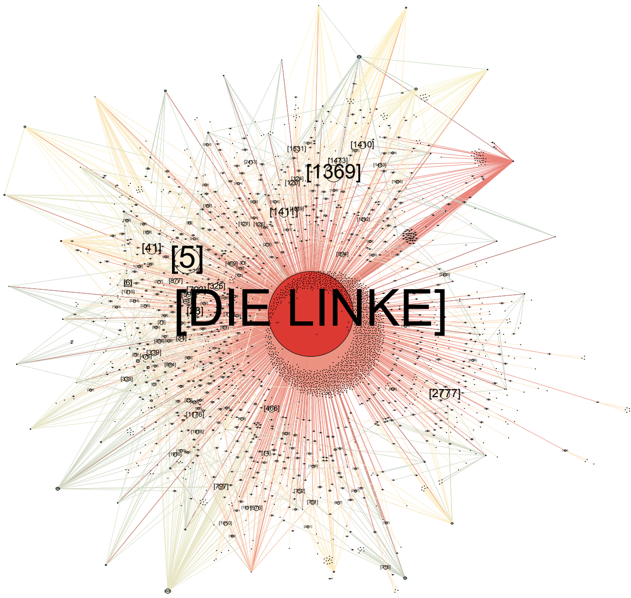
In turn one can interpret the cumulative discrepancy $\sum_i^N |CR(L_i) - LR(L_i)|$ as a function of network regularity.

VI. Conclusion and Future Work

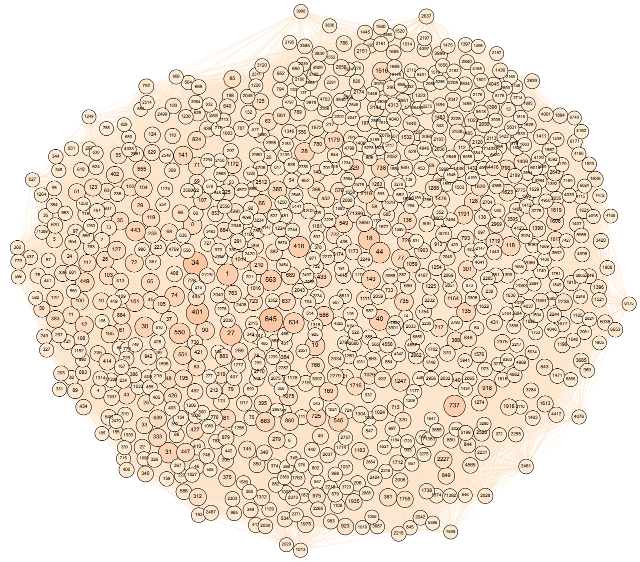
The analysis of social networks, and in particular the finding of influential and opinion-influencing profiles, is of great interest in forensic research for a variety of reasons. In the present study, it was shown that the usual centrality-based approaches, and in particular the LeaderRank, produce erronous results in star-like networks, such as Facebook pages of parties. Furthermore, LeaderRank skewness was presented as an appropriate measure to quantify the degree of distortion of a network or in other words its proximity to a star-shaped topology. Finally, the CompetenceRank was introduced as a measure that provided better results that the popular LeaderRank for the data used in the study.

In following studies, it would be interesting to analyze the observed phenomena in more fine-grained times pans as well as over a longer period of time. Additionally, it is necessary to take more and different network topologies into account. Furthermore, it was noticed that the texts in the data used were surprisingly well written. This provides an opportunity to conduct further textual analyses especially to answer the question whether there is a correlation between topics and opinion leaders and if so, how both develop over time.

 ISBN: 978-1-61208-654-5



(a) The network of the facebook page "DIE LINKE" of January 2017.



(b) Part of the *Epinions* social network [19] (filtered by k-core $\geq$ 80).

Figure 3. Comparison of two different network topologies.

REFERENCES

[1] M. Spranger, F. Heinke, S. Grunert, and D. Labudde, "Towards predictive policing: Knowledge-based monitoring of social networks," in The Fifth International Conference on Advances in Information Mining and Management (IMMM 2015), 2015, pp. 39 – 40.

[2] M. Spranger, H. Siewerts, J. Hampl, F. Heinke, and D. Labudde, "SoNA: A Knowledge-based Social Network Analysis Framework for Predictive Policing," International Journal On Advances in Intelligent Systems, vol. 10, no. 3 & 4, 2017, pp. 147 – 156.

[3] M. Spranger, S. Becker, F. Heinke, H. Siewerts, and D. Labudde, "The infiltration game: Artificial immune system for the exploitation of crime relevant information in social networks," in Proc. Seventh International Conference on Advances in Information Management and Mining (IMMM), IARIA. ThinkMind Library, 2017, pp. 24–27.

[4] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," Comput. Netw. ISDN Syst., vol. 30, no. 1-7, Apr. 1998, pp. 107–117.

[5] L. Lü, Y.-C. Zhang, C. H. Yeung, and T. Zhou, "Leaders in social networks, the delicious case," PloS one, vol. 6, no. 6, 2011, p. e21202.

[6] Q. Li, T. Zhou, L. Lü, and D. Chen, "Identifying influential spreaders by weighted leaderrank," Physica A: Statistical Mechanics and its Applications, vol. 404, no. Supplement C, 2014, pp. 47 – 55.

[7] E. Katz, "The two-step flow of communication: An up-to-date report on an hypothesis," Public Opinion Quarterly, vol. 21, no. 1, Anniversary Issue Devoted to Twenty Years of Public Opinion Research, 1957, p. 61.

[8] E. M. Rogers, Diffusion of innovations. New York: The Free Press, 1962.

[9] B. Huang, G. Yu, and H. R. Karimi, "The finding and dynamic detection of opinion leaders in social network," Mathematical Problems in Engineering, vol. 2014, 2014, pp. 1–7.

[10] P. Parau, C. Lemnaru, M. Dinsoreanu, and R. Potolea, "Opinion leader detection." in Sentiment analysis in social networks, F. A. Pozzi, E. Fersini, E. Messina, and B. Liu, Eds., 2016, pp. 157–170.

[11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," J. Mach. Learn. Res., vol. 3, Mar. 2003, pp. 993–1022.

[12] X. Song, Y. Chi, K. Hino, and B. Tseng, "Identifying opinion leaders in the blogosphere," in Proceedings of the sixteenth ACM conference on Conference on information and knowledge management - CIKM '07, M. J. Silva, A. O. Falcão, A. A. F. Laender, R. Baeza-Yates, D. L. McGuinness, B. Olstad, and Ø. H. Olsen, Eds. New York, New York, USA: ACM Press, 2007, pp. 971 – 974.

[13] B. Amor et al., "Community detection and role identification in directed networks: Understanding the twitter network of the care.data debate," CoRR, vol. abs/1508.03165, 2015.

[14] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, Za&iuml and O. R. ane, Eds. New York, NY: ACM, 2002, p. 61.

[15] C. Egger, "Identifying key opinion leaders in social networks: An approach to use instagram data to rate and identify key opinion leader for a specific business field," Master Thesis, TH Köln - University of Applied Sciences, Köln, 2016.

[16] M. Z. Shafiq, M. U. Ilyas, A. X. Liu, and H. Radha, "Identifying leaders and followers in online social networks," IEEE Journal on Selected Areas in Communications, vol. 31, no. 9, 2013, pp. 618–628.

[17] Z. H. Zhang, G. P. Jiang, Y. R. Song, L. L. Xia, and Q. Chen, "An improved weighted leaderrank algorithm for identifying influential spreaders in complex networks," in 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), vol. 1, July 2017, pp. 748–751.

[18] F. Passerini and S. Severini, "Quantifying complexity in networks: The von neumann entropy," Int. J. Agent Technol. Syst., vol. 1, no. 4, Oct. 2009, pp. 58–67.

[19] M. Richardson, R. Agrawal, and P. Domingos, "Trust management for the semantic web," in The Semantic Web - ISWC 2003, D. Fensel, K. Sycara, and J. Mylopoulos, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 351–368.