Relating Real and Synthetic Social Networks Through Centrality Measures

Maria J. Blesa
Computer Science Department, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

Mihail Eduard Popa
Barcelona School of Informatics, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

Maria Serna
Computer Science Department, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
Institute of Mathematics (IMTech), Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

Abstract

We perform here a comparative study on the behaviour of real and synthetic social networks with respect to a selection of nine centrality measures. Some of them are topology based (degree, closeness, betweenness), while others consider the relevance of the actors within the network (Katz, PageRank) or their ability to spread influence through it (Independent Cascade rank, Linear Threshold Rank). We run different experiments on synthetic social networks, with 1K, 10K, and 100K nodes, generated according to the Gaussian Random partition model, the stochastic block model, the LFR benchmark graph model and hyperbolic geometric graphs model. Some real social networks are also considered, with the aim of discovering how do they relate to the synthetic models in terms of centrality. Apart from usual statistical measures, we perform a correlation analysis between all the nine measures. Our results indicate that, in general, the correlation matrices of the different models scale nicely with size. Moreover, the correlation plots distinguish four categories that classify most of the real networks studied here. Those categories have a clear correspondence with particular configurations of the models for synthetic networks.

2012 ACM Subject Classification Networks → Network algorithms; Networks → Network dynamics

Keywords and phrases centrality measures, influence spread models, synthetic social networks

Digital Object Identifier 10.4230/LIPIcs.SEA.2022.7

Supplementary Material *Other (Experimental Results)*: `https://www.cs.upc.edu/ mjblesa/centrality/syntheticGraphs/`

Funding M. Blesa and M. Serna acknowledge support by MICIN/AEI/10.13039/501100011033 under grant PID2020-112581GB-C21 (MOTION). M.E. Popa was funded by the Ministry of Education and Vocational Training (MEFP) with a student grant (Beca de colaboración, call 2020-2021).

1 Introduction

Nowadays, social media are more and more integrated in our daily lives leading to the emergence of varied and complex social networks. One of the main research questions is to understand the relevant characteristics of those huge networks. In network analysis, indicators of centrality identify the most important vertices within a graph with respect to some particular characteristic. Centrality concepts were first developed in social network analysis, and many of the terms used to measure them reflect that sociological origin [17]. They should not be confused with node influence metrics, which seek to quantify the influence of every node in the network. Traditional measures are *degree*, *closeness* and *betweenness* which are topology dependant. Other well-known centrality measures are the Katz Rank [8] and the PageRank [19]. Two new measures have been introduced in an attempt to measure centrality with respect to influence spreading, the Independent Cascade (ICR) [9] and the Linear Threshold Rank (LTR) [4].

Many efforts have been devoted to understand the relationship among different centrality measures on real networks (see for example [26, 22, 18] and references herein). [18] show that network topology determines the correlation pattern among several measures and the combination of several centrality measures can help in the interpretation of the roles that a node or a group of nodes play in the network. In all these studies the evaluation has been performed in a selection of real networks. Our objective is to understand whether the parameters and correlation patterns among a small number of centrality measure can identify a family of synthetic social networks. Eventually, we would like to use such patterns to associate a correct model to real networks. Recall that, after a surge in interest in network structure among mathematicians and physicists, a body of research has been devoted to modeling networks either analytically or numerically. We focus our attention on random models generating synthetic networks from a priori knowledge on its communities. In particular, we generate graphs according to the *Gaussian Random partition* model [25], the *stochastic block* model [20], and the *Lancichinetti–Fortunato–Radicchi (LFR) benchmark* [13]. These models are accepted to be good generators for benchmarks in community detection [3]. We complement the study analyzing *hyperbolic geometric graphs* [12], known as graphs showing properties expected in large real network graphs, and the behaviour of the centrality on some real social networks.

We experimentally evaluate the behaviour of the considered centrality measures on the selected models of synthetic networks and perform a stochastic comparison among them. To compare the different centrality measures, we have extracted three statistics measuring diversity: the standard deviation, the number of different values, and the Gini coefficient. The second component is a correlation analysis. We use the Kendall [11] and Spearman [24] coefficients. Our results show, as expected, that the considered models of synthetic social networks perform differently with respect to the centrality measures. Besides this, our results indicate that, in general, the correlation matrices on the different models scale nicely with size. Furthermore, we observe that the correlation plots distinguish different graph families. From the correlation plots, we have been able to obtain four categories classifying most of the real networks studied here. Furthermore, such categories are identified with submodels of the synthetic networks. The unclassified networks are very small or a particular structure which hints to another type of generator will be needed for those networks.

2 Centrality Measures

We outline the centrality measures used in the paper. Some of them are based on the topological properties of the nodes, some others take into account the relevance of the actors, while other centrality measures quantify the influence exerted by the actors on the network in terms of diffusion power. Given a directed graph $G = (V, E)$, where $|V| = n$, we consider:

2.1 Topology based

Degree. This is one of the simplest centrality measures and simply consists on assigning the centrality based of the degree of the node. For every node $i \in V$, the degree centrality of i is the degree itself of i (δ_i), normalized by the number of nodes minus one, i.e.,

$$\mathsf{Deg}(i) = \frac{\delta_i}{n-1}.$$

Closeness. For the closeness measure, the distance to other nodes is considered. Intuitively, the more central a node is, the closer it is to all other nodes. For every node $i \in V$, the closeness of i is calculated as the normalized reciprocal of the sum of the length of the shortest path distances between i and all other nodes $j \in V \setminus \{i\}$ in the graph.

$$\mathsf{Clsn}(i) = \frac{1/\sum_j d(j,i)}{(n-1)} = \frac{n-1}{\sum_j d(j,i)}.$$

In order to be able to compute the closeness measure also for the nodes of networks which are not strongly connected, we will consider the adaptation in [27], defined as follows:

$$\mathsf{Clsn}(i) = \frac{J_i/(n-1)}{\sum_j d(j,i)/J_i}.$$

where J_i is the number of actors in the influence range of actor i, i.e., the number of actor who are reachable from i.

Betweenness. In the betweenness centrality, a node is more important if it belongs to the shortest path between any pair of nodes in the graph [5]. Given $G = (V, E)$, for every node $i \in V$, we define the betweenness as the sum $\forall s \forall t \in V$ of the proportion of $\sigma_{st}(i)$ (the shortest paths between s, t that go through i) with respect to σ_{st} (all the shortest path between s and t), i.e.,

$$\mathsf{Btwn}(i) = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}.$$

2.2 Relevance based

Katz. The Katz centrality [8] is a generalization of degree centrality and it can also be viewed as a variant of eigenvector centrality. While degree centrality measures the number of direct neighbors, the Katz centrality measures the number of all nodes that can be connected through a path, while the contributions of distant nodes are penalized. It is based on the idea that an actor is important if it is linked to other important actors or if it is highly linked. It overcomes the limitations of the eigenvector centrality when the graph has nodes that reach strongly connected components, but those connected components do not reach the node, which may occur in social networks.

Let A be the adjacency matrix of the directed graph $G = (V, E)$ (i.e., $a_{ij} = 1$ if there is an edge between i and j, and a $a_{ij} = 0$ otherwise). Let β be a constant independent from the structure of the social network and $\alpha \in [0, \ldots, {\lambda_{max}}^{-1}]$ is the damping factor, being λ_{max} the highest eigenvalue in A. Then the $Katz(i)$ is defined as

$$\mathsf{Ktz}(i) = \alpha \sum_{j \in V} a_{ji}\, \mathsf{Ktz}(j) + \beta.$$

PageRank. One of the most popular centrality measures is the PageRank [19], that Google uses to assign importance to web pages. A web page is important if other important web pages point to it. It uses a parameter $\alpha \in (0, 1]$, that represents the probability that a user keeps jumping from a web page to another through the links that are between them (and thus, $1 - \alpha$ represents the probability that the user goes to a random web page). Let A be the adjacency matrix of a directed graph $G = (V, E)$ (i.e., $a_{ij} = 1$ if there is an edge between i y j, and a $a_{ij} = 0$ otherwise), and let $\delta^+(i)$ be the out degree of $i \in V$. The PageRank (PR) of i is given by

$$\mathsf{PgR}(i) = (1 - \alpha) + \alpha \sum_{j \in V} \frac{a_{ji}\,\mathsf{PgR}(j)}{\delta^+(j)}.$$

2.3 Influence based

Perhaps the two most prevalent diffusion models in computer science are the *Independent Cascade* model [7] and the *Linear Threshold* model [9, 23]. Based on them, the corresponding influence-based centrality measures are defined:

Independent Cascade Rank. The Independent Cascade Rank [10] is an influence-based centrality measure based on the Independent Cascade Model (ICM) [7], which is a stochastic model that was initially proposed in the context of marketing. It is based on the assumption that whenever a node is activated, it will (stochastically) do attempt to activate each actor he targets. Given an activated node $i \in V$, any neighbor j such that $(i, j) \in E$ will be activated with a probability p_{ij}. When a new actor is activated, the process is repeated for this actor. The whole process ends when there are no active nodes with a new chance to spread its influence.

Given an initial node $u \in V$ and a probability $p \in [0, 1]$ (where $\forall (i, j) \in E : p_{ij} = p$), the expected influence spread of u is denoted by $F'(u, p)$ and comprises the set of activated nodes under the ICM influence model, starting from the initial node u. Then, the Independent Cascade Rank of a node $u \in V$ is then defined as

$$\mathsf{ICR}(u, p) = \frac{|F'(u, p)|}{\max_{v \in V}\{|F'(v, p)|\}}.$$

Linear Threshold Rank. The Linear Threshold Rank [4] is also an influence-based centrality measure, based on the Linear Threshold Model (LTM) [9]. Every node has an influence threshold, which represents the resistance of this node to be influenced by others. Every edge (u, v) also has a weight representing the influence that node u has over node v.

The influence algorithm starts with an initial predefined set of activated nodes. At every iteration, the active nodes will influence their neighbors. When the total influence that a node i receives exceeds its influence threshold θ_i, then this node will become active and join the set of active nodes. As long as new nodes join the set of active nodes, the spread of influence is still on progress. The algorithm stops when the set of active nodes converges, i.e., when no new nodes are influenced. In order to formally define the Linear Threshold Rank, we need to introduce the following concepts:

▶ **Definition 1.** *An influence graph is a tuple (G, w, θ), where $G = (V, E)$ is a directed graph made by a set of actors V and a set of relations $E, w : E \to \mathbb{Z}$ is a weight function that assigns a weight to each edge, representing the influence of one node to the other, and $\theta : V \to \mathbb{N}$ is a labeling function that quantifies how resistant to influence every node is.*

▶ **Definition 2.** *Given an influence graph (G, w, θ) and an initial active set $X \subseteq V$, $F_t(X) \subseteq V$ denotes the set of activated nodes at the t-th iteration starting with X as kernel.*

At the first step, $t = 0$ only the nodes in X are active, which means that $F_0(X) = X$. At the $t+1$ iteration, a node i will be activated if, and only if, the sum of all the weights of the active nodes incident to i is higher than the resistance (or influence) threshold of i, i.e.,

$$\sum_{j \subseteq F_t(X)} w_{ij} \geq \theta_i,$$

Observe that the process is monotonic, therefore it stops after at most $n = |V|$ steps.

▶ **Definition 3.** *Let $k = min\ \{t \in \mathbb{N} \mid F_t(X) = 0\}$, where $k \leq n$. The expansion of $X \subseteq V$ on an influence graph $(G = (E, V), w, \theta)$, is defined as $F(X) = \bigcup_{t=0}^{k} F_t(X)$.*

Given an influence graph $(G = (V, E), w, \theta)$, the Linear Threshold Rank of a node $i \in V$ is given by

$$\mathsf{LTR}(i) = \frac{|F(\{i\} \cup \mathcal{N}(i))|}{|V|}, \text{ where } \mathcal{N}(u) = \{v \mid (u, v) \in E \vee (v, u) \in E\}.$$

Forward and Backward Linear Threshold Rank. The Forward and the Backward Linear Threshold Ranks [1] are centrality measures similar to the LTR, but with a different initial set of activated nodes. Given an influence graph $(G = (V, E), w, f)$, the Forward Linear Threshold Rank and the Backward Linear Threshold Rank of a node $i \in V$ is given, respectively, by

$$\mathsf{FwLTR}(i) = \frac{|F(\{i\} \cup \mathcal{N}^+(i))|}{|V|} \qquad \text{and} \qquad \mathsf{BwLTR}(i) = \frac{|F(\{i\} \cup \mathcal{N}^-(i))|}{|V|}$$

where $\mathcal{N}^+(i) = \{j \in V \mid (i, j) \in E\}$ and $\mathcal{N}^-(i) = \{j \in V \mid (j, i) \in E\}$.

3 Social Networks

We describe the characteristics of the synthetic networks considered in this work, as well as the real social networks chosen for our experiments. The structural characteristics of the networks are described by seven common attributes: the number of vertices, the number of edges, whether the graph is weighted, whether the graph is directed, the average clustering coefficient, and the size of the main core.

The *average clustering coefficient* (ACC) is the average of the local clustering coefficients in the graph. The local clustering coefficient C_i of a node i is the number of triangles T_i in which the node participates normalized by the maximum number of triangles that the node could participate in.

$$\text{ACC} = \frac{1}{n} \sum_{i=1}^{n} C_i, \quad \text{where } \ C_i = \frac{T_i}{\delta_i(\delta_i - 1)}$$

where δ_i is the degree of the node i, and $n = |V|$. Given a graph G and $k \in \mathbb{Z}^+$, a *k-core* is the maximal induced subgraph of G where every node has at least degree k. The *main core* is a k-core of G with the highest k.

3.1 Synthetic Social Networks

Concerning the models for synthetic social networks, we have considered four different models: Gaussian Random Partition Graphs, the Stochastic Block Model, LFR Benchmark Graphs and Hyperbolic Geometric Graphs.

3.1.1 Gaussian Random Partition Graph (GRP)

The process to create a Gaussian Random Partition Graph [25] starts by creating k partitions of different size. Those sizes will be taken from a normal distribution $\mathcal{N}(\mu, \sigma^2)$. Two nodes from the same partition are connected with probability p_{in}, while two nodes from two different partitions will be connected with probability p_{out}. To generate this type of graph we will be using the implementation from NetworkX [14], which lets us assign values for the following parameters:

- n: the number of nodes of the network,
- μ: mean of the sizes of the partition in the graph,
- σ: variance of the sizes of the partition in the graph,
- p_{in}: probability of generating a intracluster edge,
- p_{out}: probability of generating a intercluster edge,
- dir: whether or not the graph is directed.

We created five different types of graphs, that we denote as GRPa, GRPb, GRPc, GRPd and GRPe (see Table 1). They consider different size and variance of partitions, and also different probabilities for creating intracluster and intercluster edges. All categories are directed, except for GRPd. The choice for these parameters is based on the ones proposed in [25], but conveniently adapted to represent bigger meaningful social networks.

3.1.2 Stochastic Block Model (SBM)

The construction of a Stochastic Block Model Graph [20] starts by partitioning the nodes of the network into blocks of arbitrary sizes. Secondly, edges are placed between pairs of nodes independently, with a probability that depends on the blocks, i.e., the probability to create an edge (u, v) depends on the probability of connection defined between the cluster of u and the cluster of v.

To generate this type of graph we will be using the implementation from NetworkX [16], which lets us assign values for the following parameters:

- n: the approximate number of nodes of the network,
- k: the number of blocks within the network,
- $S = \{s_1, \ldots, s_k\}$: the list of block sizes, where s_i denotes the number of nodes in the block i.
- $P \in k^2$: a probability matrix, where p_{ij} is the probability of creating an (intercluster) edge between a node in cluster i and a node in cluster j. Observe that p_{ii} is then the probability of an intracluster edge within block i.

We created five different types of graphs, that we denote as SBMa, SBMb, SBMc, SBMd and SBMe (see Table 2). The sizes of the blocks are created according to different statistical distributions. We use exponential distributions for all the types of graphs, except for the SMBb, where a normal distribution is used. All categories are directed.

We can observe that all the networks have the order of $\theta(\sqrt{n})$ number of clusters, except for SBMd, which has less clusters but of bigger size. In general, we wanted to work with very different cluster sizes. For that reason, in most of the cases we used the exponential distribution to generate S. For the SBMb we used the normal distribution instead. We want to see whether big differences on the size of the blocks affect the final result.

Table 1 Parameters for the Gaussian Random Partition Graphs. For each type, three different graph sizes are considered: $n = 1000$, $n = 10000$ and $n = 100000$. For the probabilities, $f(n) = \frac{\log(n)}{\mu}$ and $g(n) = \frac{\log(n)}{n-\mu}$.

Name	μ	σ	p_{in}	p_{out}	*dir*
GRPa	$n/10\sqrt{n}$	4	$3/4\, f(n)$	$1/4\, g(n)$	True
GRPb	$n/10\sqrt{n}$	2	$3/4\, f(n)$	$1/4\, g(n)$	True
GRPc	$n/10\sqrt{n}$	4	$1/2\, f(n)$	$1/2\, g(n)$	True
GRPd	$n/10\sqrt{n}$	4	$3/4\, f(n)$	$1/4\, g(n)$	False
GRPe	$n/\sqrt{n}$	4	$3/4\, f(n)$	$1/4\, g(n)$	True

Table 2 Parameters for the Stochastic Block Model Graphs. For each type, three different graph sizes are considered: $n = 1000$, $n = 10000$ and $n = 100000$. $S \sim Exp(\lambda)$ is a sample from an exponential distribution with rate λ, and $S \sim \mathcal{N}(\mu, \sigma^2)$ is a normal distribution with mean μ and standard deviation σ^2. For the probabilities, $f(n) = \log(n)/\overline{S}$ and $g(n) = \log(n)/(n - \overline{S})$, where $\overline{S}$ represents the mean size of the blocks.

Name	k	S	p_{ii}	p_{ij}
SBMa	$10\sqrt{n}$	$Exp(10/\sqrt{n})$	$3/4\, f(n)$	$1/4\, g(n)$
SBMb	$10\sqrt{n}$	$\mathcal{N}(k/100, k/1000)$	$3/4\, f(n)$	$1/4\, g(n)$
SBMc	$10\sqrt{n}$	$Exp(10/\sqrt{n})$	$1/2\, f(n)$	$1/2\, g(n)$
SBMd	$\sqrt{n}$	$Exp(1/\sqrt{n})$	$3/4\, f(n)$	$1/4\, g(n)$
SBMe	$10\sqrt{n}$	$Exp(10/\sqrt{n})$	$3/4\, \frac{\log(n)}{s_i}$	$1/4 \frac{\log(n)}{n-s_i}$

3.1.3 LFR Benchmark Graph

The LFR Benchmark [13] is a model for graph generation more complex than GRP and SBM are. Consequently, it allows to create artificial networks that are significantly more similar to real ones. In a very summarized way, the algorithm starts finding a power law distribution for the degree of the nodes. Every node will have a proportion μ of its connections to nodes belonging to other communities (intercluster), whereas the remaining $(1 - \mu)$ proportion of its edges will be attached to nodes in same the community (intracluster). This leads to the emergence of communities of different sizes (following a power law distribution as well). Each node will be randomly assigned to one community following the constraint imposed by μ.

To generate this type of graph we will be using the implementation from NetworkX [15], which lets us assign values for a lot of different parameters. After a deep study of them, where we detected some incompatibilities, the following parameters where identified as the most relevant and worth to play with:

- n: the number of nodes that our social network will have,
- τ_1: exponent of the power law distribution for the node degree distribution,
- τ_2: exponent of the power law distribution for the community size distribution,
- μ: proportion of intracluster edges for each node,
- max_c: maximum community size in the graph,
- min_c: minimum community size in the graph,
- max_d: maximum node degree in the graph,
- avg_d: mean node degree in the graph.

We create five different types of graphs, that we denote as LFRa, LFRb, LFRc, LFRd and LFRe (see Table 3). We have fixed the parameters τ_1 and τ_2 to the values proposed in [13]: the adequate values to represent social networks oscillate between $2 \leq \tau_1 \leq 3$ and $1 \leq \tau_2 \leq 2$. We do not work with $\tau_2 = 1$ because the generator implemented in NetworkX demands that $\tau_2 > 1$. LFRc builds bigger communities than the rest of the graphs, and in LFRe the proportion of intracluster edges is greater than the usual 0.2.

Table 3 Parameters for the LFR Benchmark Graphs. For each type, three different graph sizes are considered: $n = 1000$, $n = 10000$ and $n = 100000$.

Name	τ_1	τ_2	μ	max_c	min_c	max_d	avg_d
LFRa	2	1.1	0.2	0.05 n	$max_c/100$	0.05 n	$5/4\ \log(n)$
LFRb	2	2	0.2	0.05 n	$max_c/100$	0.05 n	$5/4\ \log(n)$
LFRc	2	1.1	0.2	0.05 n	$max_c/10$	0.05 n	$5/4\ \log(n)$
LFRd	3	2	0.2	0.05 n	$max_c/100$	0.05 n	$5/4\ \log(n)$
LFRe	2	1.1	0.35	0.05 n	$max_c/100$	0.05 n	$5/4\ \log(n)$

Table 4 Parameters for the Hyperbolic Geometric Graphs. For each type, three different graph sizes are considered: $n = 1000$, $n = 10000$ and $n = 100000$.

Name	k	γ	t	z
HYPa	$\log(n)$	2	0	1
HYPb	$\log(n)$	2	2	1
HYPc	$\log(n)$	3	0	1
HYPd	$\log(n)$	3	0.5	1
HYPe	$\log(n)$	∞	0.5	1

Table 5 Real data sets considered in this work. Shadowed rows state for directed networks. ACC = Average Clustering Coefficient, MC = size of the main core. The diameter is ∞ when the graph is not connected (or not strongly connected in the case of digraphs); in these cases, the diameter of the biggest connected component is provided. The subscript (w) indicates edge weighted networks.

Networks	n	m	ACC	Diameter	MC
Dining Table $_{(w)}$	26	52	0.1178	∞ (6)	20
Dolphins	62	159	0.2590	8	36
Human Brain $_{(w)}$	480	1000	0.3004	∞ (20)	11
ArXiv	5242	14496	0.5296	∞ (17)	44
Wikipedia	7115	103689	0.1409	7	336
Caida $_{(w)}$	26475	106762	0.2082	17	50
ENRON	36692	183831	0.4970	11	275
Gnutella	62586	147892	0.0055	11	1004
Epinions	75879	508837	0.1378	14	422
Higgs $_{(w)}$	256491	328132	0.0156	19	10
Amazon	334863	925872	0.3967	44	497
Texas	1379917	1921660	0.0470	1054	1579

The parameters max_c and max_d are both fixed to 0.05 n for all the networks. For lower values, the resulting graphs are not good representations for social networks. For bigger values, the generator takes an enormous amount of time to converge (even with a very small number of vertices) or even does not converge at all. In the same sense, another problematic parameter was avg_d. When fixed to $avg_d = \log(n)$, the generator was no always converging and thus, we had to slightly increase that value. In our case, that increase implies increasing the average degree by one.

3.1.4 Hyperbolic Geometric Graph

Recent studies in graph geometry showed that many networks appearing in nature or representing societies can be modeled as geometric graphs in hyperbolic spaces [21]. Based on this fact, hyperbolic geometric graphs have started to be used as models for synthetic social networks. In our study we will be using the generator proposed in [21], which allows us to configure the following parameters:

- n: the number of nodes that our social network will have,
- k: mean node degree in the graph,
- γ: exponent of the power law distribution for the node degree distribution,
- t: temperature,
- z: square root of the curvature of the hyperbolic space.

Using these parameters, we create five different types of graphs that we denote as HYPa, HYPb, HYPc, HYPd and HYPe (see Table 4). Most of the values we work with are those suggested in [21] for generating hyperbolic geometric graphs representing social networks.

3.2 Real Social Networks

We will also be using real social networks in our experiments in order to have a point of view of what happens in reality and to see how our artificial models really compare. We consider twelve well-known real social networks, which are mostly available at the `snap.stanford.edu` and the `networkrepository.com` repositories. The main characteristics of these networks are summarized in Table 5.

4 Statistical Metrics

For analysing the results of a centrality measure on its own, three statistical metrics will be used: the number of different ranks, the standard deviations and the Gini coefficient of the distribution. The Gini coefficient [6, 2] comes originally from sociology as a measure of the inequality of populations with respect to different criteria (e.g., wealth spread), but it is lately being used as a measure for quantifying the fairness of distributions in other areas.

▶ **Definition 4.** *Given a list of values $\mathcal{X}$ of size n, the Gini coefficient of $\mathcal{X}$ is calculated as:*

$$Gini(\mathcal{X}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|}{2n\sum_{i=1}^{n} x_i}$$

As observed in other works, we will also use two well-known correlation measures for comparing the centrality ranks among themselves: the Spearman's correlation coefficient [24] and the Kendall correlation coefficient [11].

▶ **Definition 5.** *Given two lists of elements $\mathcal{X}$, $\mathcal{Y}$ both with n elements, the Spearman's rank correlation coefficient (ρ) is equal to:*

$$\rho(\mathcal{X},\mathcal{Y}) = 1 - \frac{6\sum_{i=1}^{n}(x_i - y_i)^2}{n(n^2-1)}$$

▶ **Definition 6.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be two lists of elements, then the Kendall's rank correlation coefficient (τ) is defined as:*

$$\tau(\mathcal{X},\mathcal{Y}) = \frac{n_c - n_d}{0.5n(n-1)}$$

where n_c is the number of concordant pairs between $\mathcal{X}$ and $\mathcal{Y}$, and n_d is the number of discordant pairs. A pair (i,j) is concordant if either $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$. A discordant pair is one that is not concordant.

5 Experiments and results

We have generated five different configurations for each of the four synthetic models. Then we have calculated the centrality measures and statistically analyzed the results obtained on those twenty configurations under different perspectives. In particular, the correlation analysis between the different centrality measures gives us an insight about which metrics seem to behave similarly in specific types of networks. The studies carried out can help us understand the similarities of the centrality measures under diffusion processes in different types of networks, but we do not believe that at the moment they can provide any information for their inverse use, that is, on how to generate these networks.

Undirected networks are transformed into directed ones by replacing edges into bidirectional arcs. For unweighted networks, edge weights are all fixed to 1. As in [4], we fix the threshold function θ_i of every actor i to the simple majority rule. ICR works with probability $p = 0.1$. Here we can only summarize the most relevant observations. For a detailed vision of the results and exact data on them, we point the reader to `www.cs.upc.edu/~mjblesa/centrality/syntheticGraphs/`.

5.1 Statistics results

We can observe some very clear trends when looking at the complete results of our experiments. Starting with the GRP models (see Figure 1), the deviation and the number of different ranks are the highest in GRPb and lowest in GRPd. The Gini coefficient is still high in GRPb, but it is also high in GRPc, the latter is still the lowest for some metrics such as Closeness and Pagerank. The rest of the models are quite similar to each other. Remember that GRPb is the one of the networks with the highest deviation on the sizes of the clusters, which implies more differences between clusters and thus more differences between nodes. On the other hand, GRPd is the only undirected model which could explain the low values for the standard deviation and the number of different ranks.

In the LFR models (see Figure 2), there is a clearly one model with the lowest values in every scenario and for the majority of measures, specially the influence-based ones: LFRd. These low results are caused mainly by the exponents used during the generation, being in this models the highest, especially the exponent for the degree distribution. There does not seem to be a model with clearly higher results, although LFRb does get higher values in some cases.

For the stochastic block model generator (see Figure 3), STOa, STOc, STOd have the highest values for every metric except Betweenness and Pagerank, for the standard deviation, the number of different ranks and the Gini coefficient. The models with the lowest values are STOb and STOe. Similarly to GRP, these differences occur due to the sizes of the clusters, in STOb the sizes follow a normal distribution instead of a exponential distribution, this will result in more similar clusters, so more similar nodes. In the case of STOe, although the sizes of the clusters follow an exponential distribution, the probability of creating edges inside the cluster depends on the size of the clusters, which means a node who belongs to a big cluster will have a small probability, this will imply that the number of edges will be close to a node that belongs to a small cluster but has a large probability of creating edges inside of the cluster. This phenomena balances the degree distribution of the nodes to some extent. This events can be observed specially well in the models with 100K nodes.

Finally, the hyperbolic models (see Figure 4) do not seem easy to analyse. The standard deviation is low for HYPc, HYPd and HYPe for every metric except Betweeness but there are more different ranks in these three models than in HYPa or HYPb, again with some

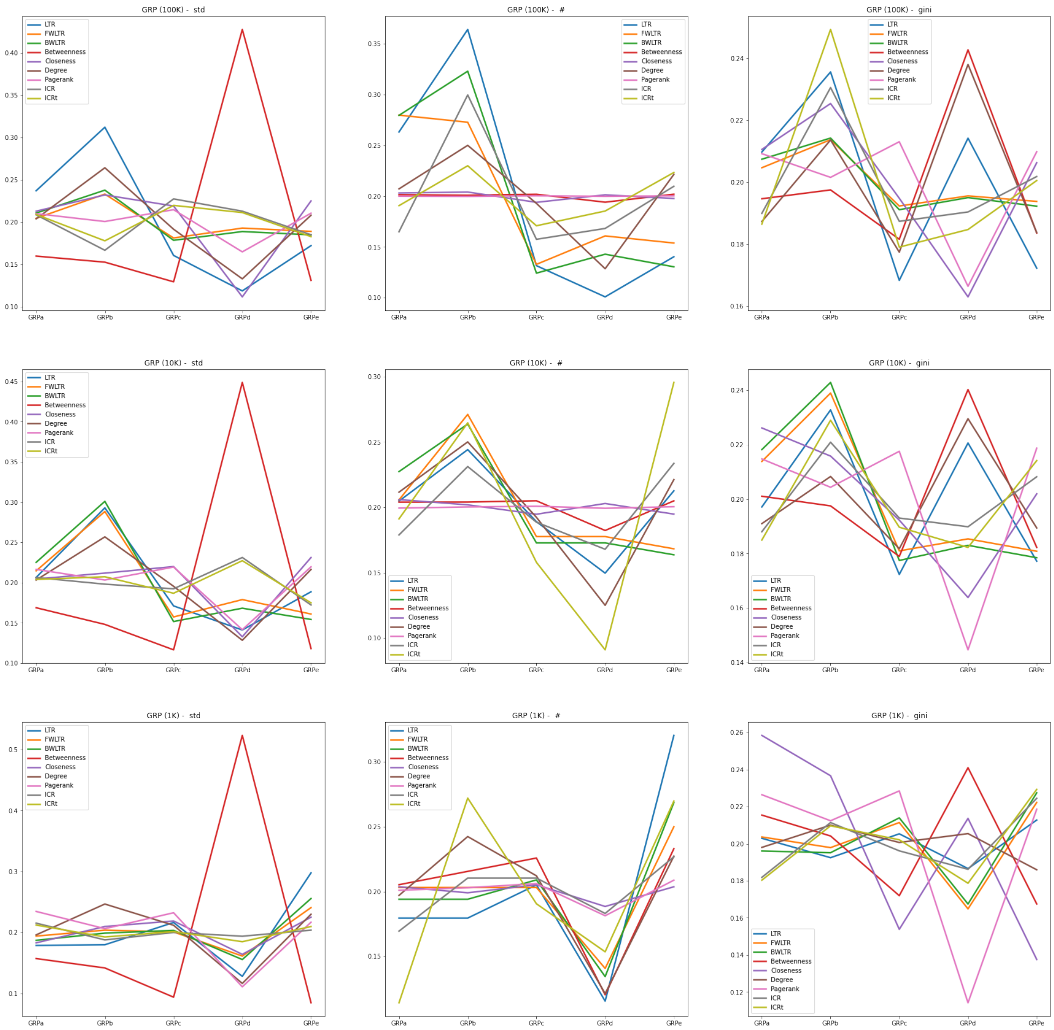

Figure 1 Statistics results for the GRP models.

exceptions like the Degree Centrality and the three centralities based on the LTM. The Gini coefficient is high in HYPa and HYPb for some metrics, but for other metrics the results in HYPc, HYPd and HYPe are higher. The only irrefutable conclusion from this network generator is that there is a clear distinction between the results for HYPa, HYPb and those for HYPc, HYPd, HYPe.

5.2 Correlation analysis

Figure 5 collects the correlation plots for the set of biggest networks (i.e., those with 100K nodes) for each of the four synthetic models under study. Figure 6 shows the correlation plots for four real social networks that represent each of the four behavioural tendencies observed. The correlation plots do not include the results of the Ktz centrality because most of the time the algorithm did not converge.

For the Gaussian graphs there is one network very different than the others, GRPd, but this happens because it is the only undirected social network, which creates the difference in the correlation patterns, having a maximum direct correlation between the three LTM

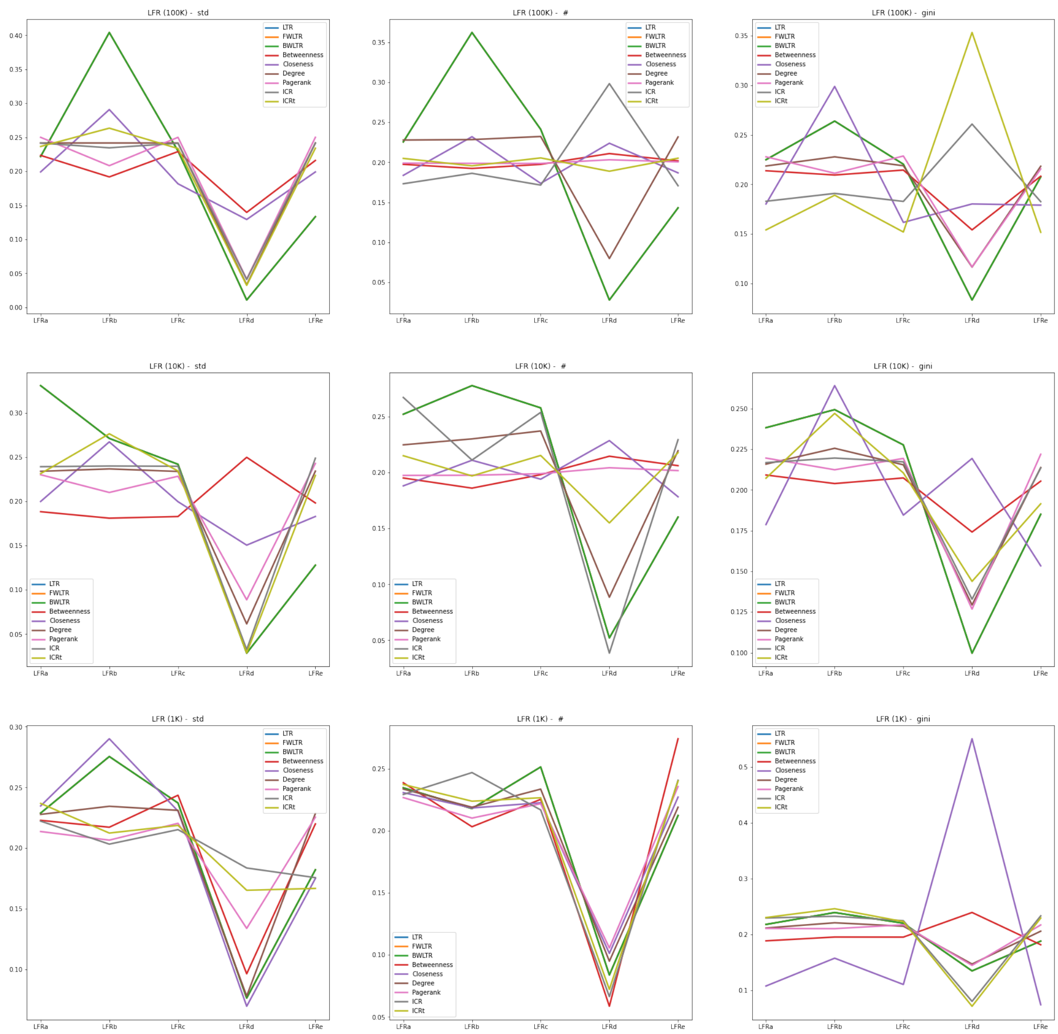

Figure 2 Statistics results for the LFR models.

based centralities. Comparing the rest of the networks we observe very similar patterns in the correlation plots, but they still can be distinguished, having more similarities between GRPa and GRPb, and between GRPc and GRPe. This differences can be seen more clearly as we decrease the number of nodes in the graph.

In the case of the LFR benchmark generator, we take into consideration that all the networks are undirected which implies that the results from LTR, FWLTR and BWLTR will be the same. The first three networks (LTRa, LTRb, LTRc) display similar patters, with low correlation between Betweenness and the LTM based metrics and high correlations between Pagerank and Degrees. One the other hand, LTRd differentiates in some aspects from these three mentioned networks, such as the high correlation between Pagerank and the LTM metrics and between Degree and the LTM metrics. Another difference is that in LFRd the correlation between Closeness and the LTM metrics is lower than the correlation between Betweenness and the LTM metrics. Finally, the LFRe network is pretty similar to the first three networks (LTRa, LTRb, LTRc) but with very subtle differences.

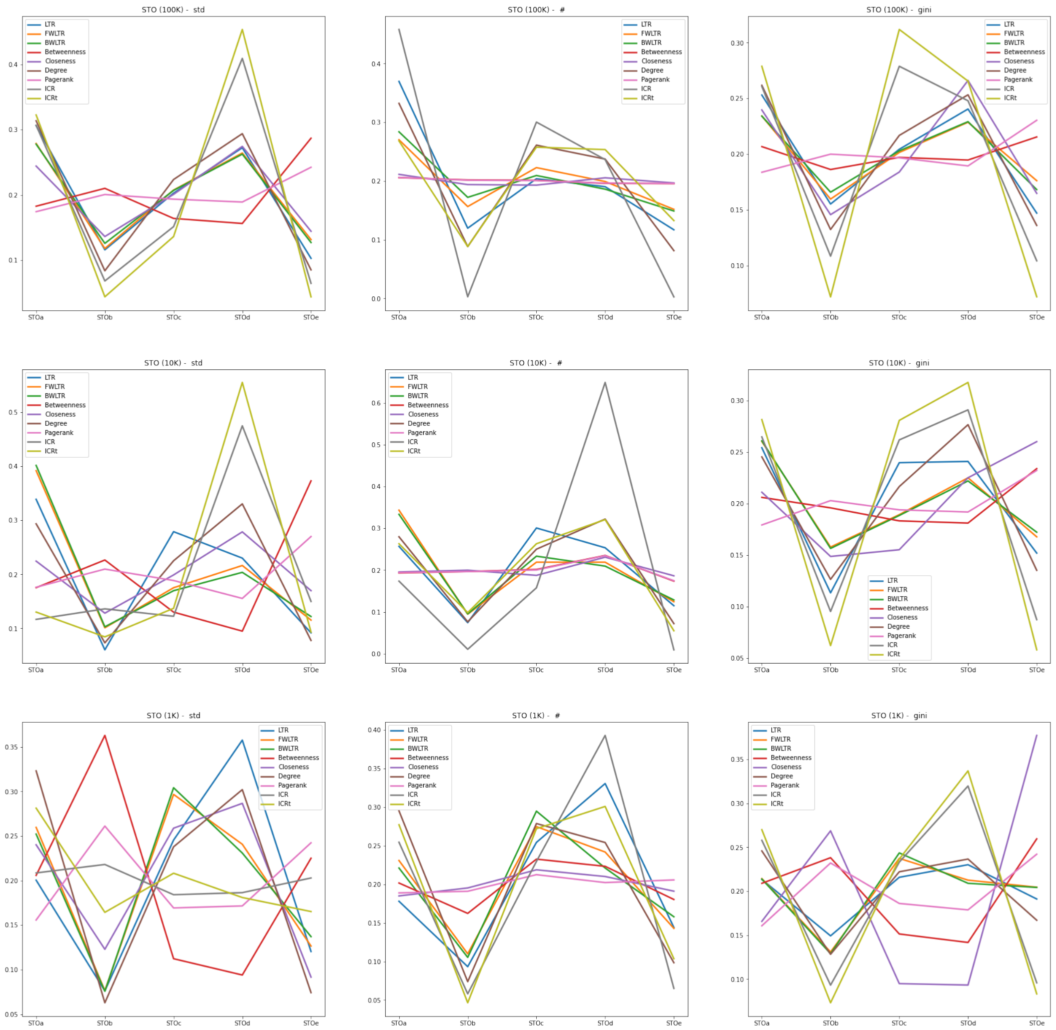

Figure 3 Statistics results for the STO models.

We can observe more differences when comparing the results from the Stochastic Block Model generator. The STOb and STOe are nearly identical, STOc has a similar pattern as the these last two, but with higher correlations between all metrics. For the other two graphs, STOa and STOd, we find most of the similarities when for big graphs with a large number of nodes, however when we compare both models with only one thousand nodes, the similarities in the patterns in the correlation plots seem to disappear, for example, STOa has very high correlations between all the LTM based metrics (LTR, FWLTR, BWLTR) but STOd does not, with very low correlations between FWLTR and BWLTR.

Finally, the networks generated in a hyperbolic space show two patterns. The first type, present in HYPa and HYPb, with high correlation between Closeness and the LTM metrics, and low correlation between the LTM metrics and almost any other metric where Betweenness and Pagerank take the lowest values, this also implies a low correlation between Closeness and, Betweenness and Pagerank. However, the other type, including HYPc, HYPd and HYPe, Closeness takes the lowest correlation with the LTM metrics. HYPc is a little different than HYPd and HYPe but the general distribution of the correlation is still the same.

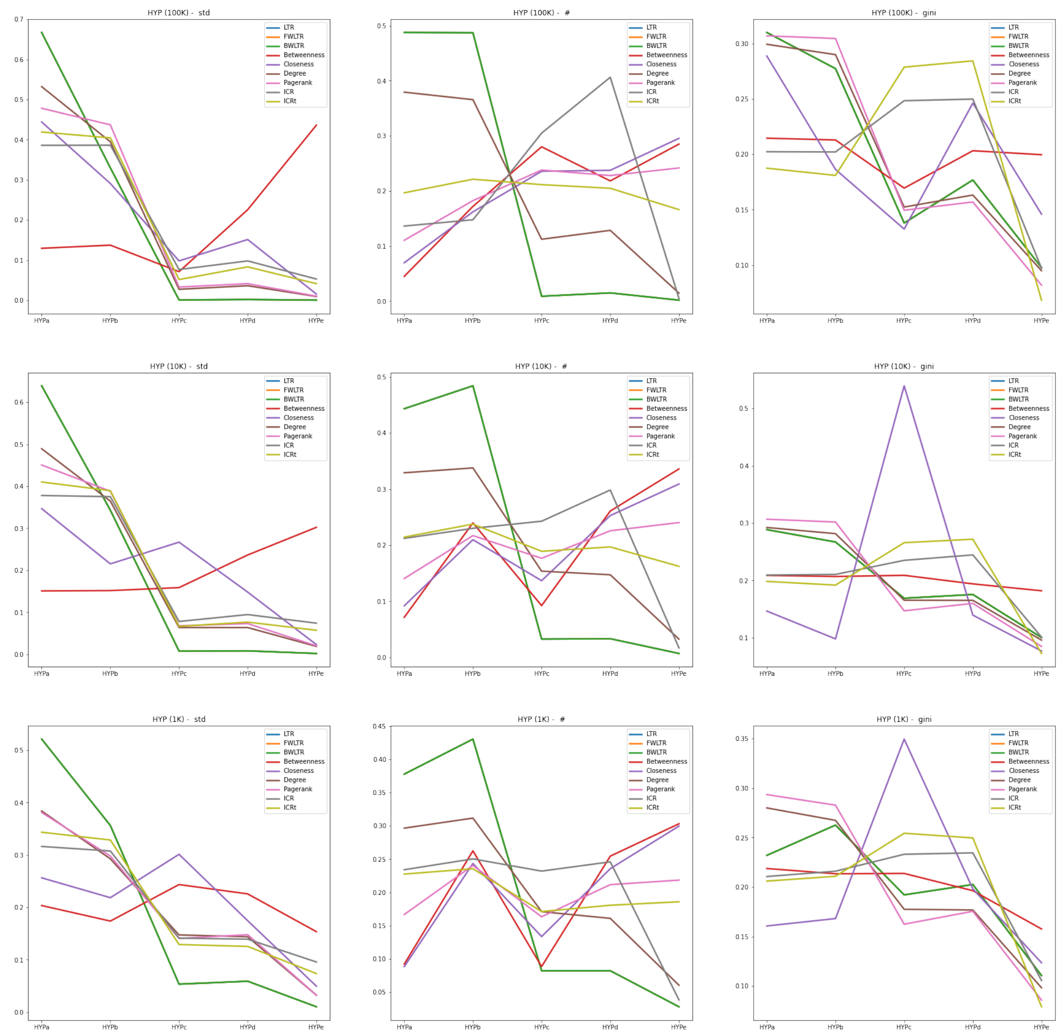

Figure 4 Statistics results for the HYP models.

5.3 Comparison with real networks

In order to extend the study of centrality measures on synthetic networks, we decided to focus on real social networks. Our aim with that was to check whether the synthetic models do really approximate real networks when centrality is concerned. We also wanted to check whether the different behavioural patterns observed in the synthetic networks would help us to classify the real networks.

Some correlation plots from real graphs look almost identical to correlation plots of synthetic networks (e.g., the Texas graph and the HYPe, the Caida graph and the LFRa). The Human Brain network is similar to most of the LFR benchmark models, but it is with a hyperbolic graph where more similarities can be found (specifically with HYPa). There are also examples of directed graphs where this also occurs, e.g., Epinions, which is pretty similar to GRPa. In most of the real social networks considered, we can observe some kind of similarity to some type of artificial network. Based on those similarities, we organize our results in four categories (two for directed graphs and two for undirected graphs). These categories are qualitative and based merely on correlations, thus describing distinguishable color patterns in the plot of the Figure 6.

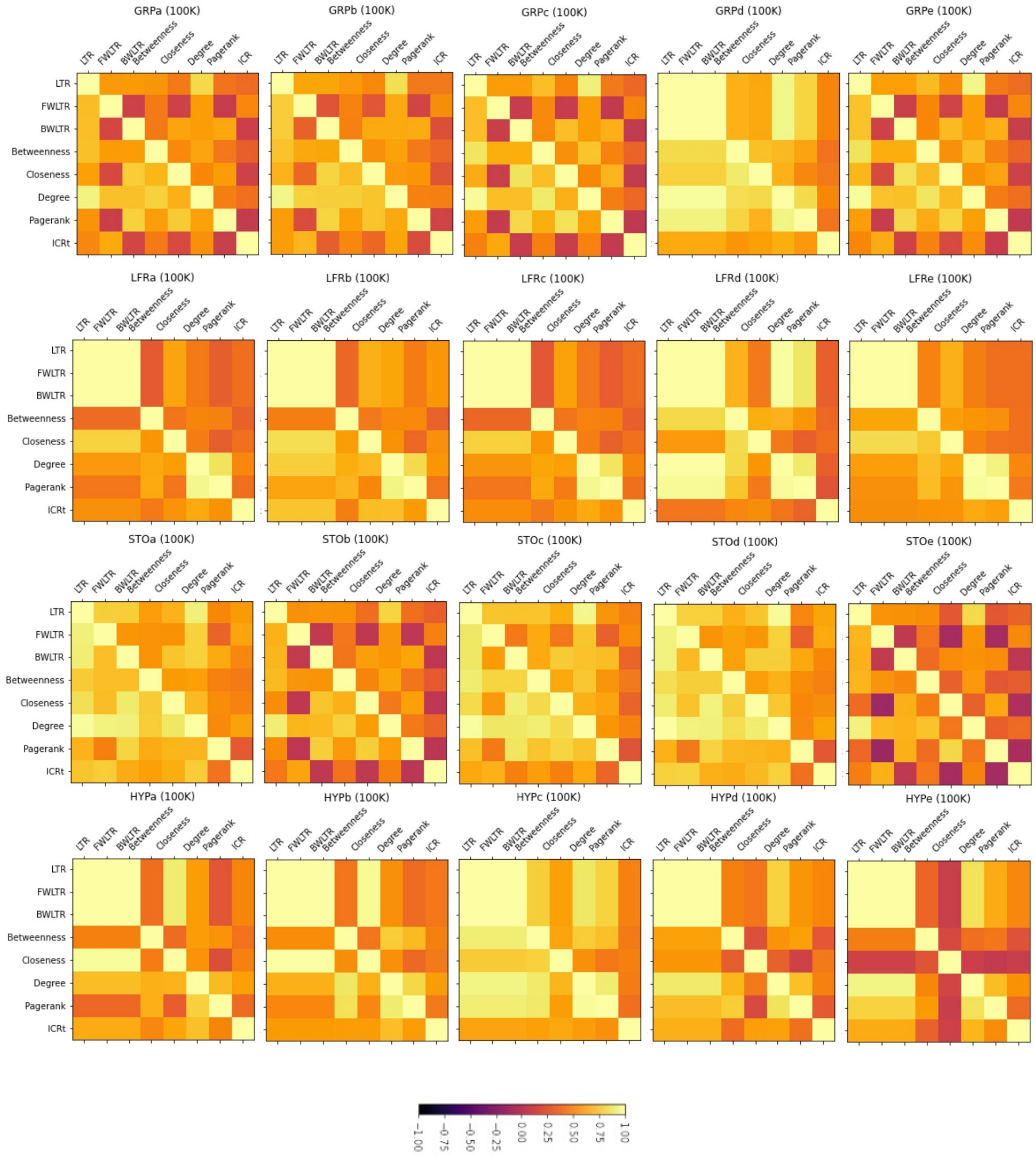

Figure 5 Heatmaps for the correlation between measures for synthetic networks with 100K nodes. Kendall coefficients are represented in the upper triangular part and Spearman in the lower one.

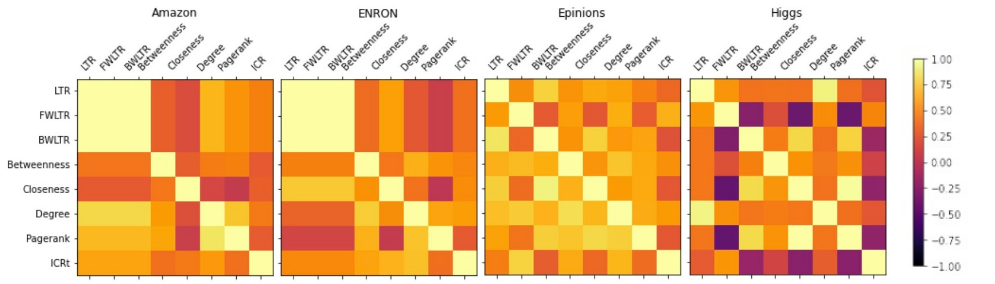

Figure 6 Heatmaps for the correlation between measures for four real social networks (Amazon, ENRON, Epinions, Higgs), which represent the four different behaviours observed experimentally. Kendall coefficients are represented in the upper triangular part and Spearman in the lower one.

The first type is for undirected graphs and includes the networks Amazon, Dolphins, Texas, and ArXiv. The synthetic networks that represent this first category are GRPd, LFRd, HYPc, HYPd, HYPe. They have low correlation between Betweenness and the LTM metrics and a very low correlation between Closeness and the LTM metrics. We find higher correlations between Degree and the LTM metrics, Pagerank and the LTM metrics, and Degree and Pagerank.

The second category is also for undirected graphs and it includes the real networks Caida, ENRON and Human Brain, and the synthetic networks LFRa, LFRb, LFRc, LFRe, HYPa and HYPb. One of the main differences between this and the previous category is the change in the correlations of Closeness and Degree: here the correlation between Degree and the LTM is high, and the correlation between Closeness and the LTM metrics is one of the highest, opposite to the first category.

We observe a third category for directed graphs, which includes the Epinions real graph and GRPa, GRPb and STOc. We found the lowest correlations when looking at FWLTR and ICRt. These two metrics have a low correlation with BWLTR, Closeness and Pagerank. The rest of the correlations are neither high nor low.

The last category that we can distinguish is also for directed graphs. In this case we have the Higgs and Wikipedia as real networks representatives, and GRPc, GRPe as synthetic graphs. The main different with the third category is that this time the low correlations of FWLTR and ICRt are much lower, with values very close to zero. In the rest of the correlations we can find low and medium correlations unlike in the last category where most of them were medium correlations.

The synthetic graphs STOb and STOe are halfway between the third and fourth category, having similarities and differences with both of them.

There are two real social networks who do not seem to belong to any of the categories mentioned, which means that they are not very similar with the synthetic networks generated in terms of the centrality measures correlation. The first example, the Dining Table network, can be easily explained. Most of the correlations given by this network have a p-value higher than 0.05 which makes most of the results not statistically significant. However, in the remaining network Gnutella, this phenomena does not occur which means than the results are valid and significant. In this case, the problem could be that the number of generators and models used is limited and does not cover all the possible networks. Another cause could be that the structure of this network is very particular and it is hard to replicate artificially with algorithms.

All the comparisons between correlations are qualitative in this work. We plan to introduce quantitative measures to be able to weight those relation, e.g. by means of similarity measures applied to the correlation matrices.

References

1 M.J. Blesa, P. García-Rodríguez, and M.Serna. Forward and backward linear threshold ranks. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 265–269. ACM, 2021. doi:10.1145/3487351.3488355.

2 L. Ceriani and P. Verme. The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality*, 10(3):421–443, 2012.

3 Hocine Cherifi, Gergely Palla, Boleslaw K. Szymanski, and Xiaoyan Lu. On community structure in complex networks: challenges and opportunities. *Applied Network Science*, 4(117):2364–8228, 2019. doi:10.1007/s41109-019-0238-9.

4 X. Molinero F. Riquelme, P. Gonzalez-Cantergiani and M. Serna. Centrality measures in social networks based on linear threshold model. *Knowledge-Based Systems*, 40:92–102, 2017. doi:10.1016/j.knosys.2017.10.029.

5 L.C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

6 C. Gini. *Variabilitàe Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche.* C. Cuppini, 1912.

7 J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development. *Technical Report, Academy of Marketing Science Review*, 2001.

8 L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18:39–43, 1953.

9 D. Kempe, J.M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. *Proceedings of the 9th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

10 D. Kempe, J.M. Kleinberg, and É. Tardos. Influential nodes in a diffusion model for social networks. In *Intl. Colloquium on Automata, Languages and Programming (ICALP)*, volume 3580, pages 1127–1138. Lecture Notes in Computer Science, 2005. doi:10.1007/11523468_91.

11 M. Kendall. A new measure of rank correlation. *Biometrika*, 30:81–93, 1938.

12 D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá. Hyperbolic geometry of complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 82((3 Pt 2):036106), 2010. doi:10.1103/PhysRevE.82.036106.

13 A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E*, 78:046110, October 2008. doi:10.1103/PhysRevE.78.046110.

14 NetworkX. Gaussian Random Partition Graph. Accessed: 2022-02. URL: https://networkx.org/documentation/stable/reference/generated/networkx.generators.community.gaussian_random_partition_graph.html.

15 NetworkX. LFR Benchmark Graph. Accessed: 2022-02. URL: https://networkx.org/documentation/stable/reference/generated/networkx.generators.community.LFR_benchmark_graph.html.

16 NetworkX. Stochastic Block Model. Accessed: 2022-02. URL: https://networkx.org/documentation/stable/reference/generated/networkx.generators.community.stochastic_block_model.html.

17 Mark E.J. Newman. *Networks: An introduction.* Oxford University Press, 2010. doi:10.1093/acprof:oso/9780199206650.001.0001.

18 S. Oldham, B. Fulcher, L. Parkes, A. Arnatkevičiūté, C. Suo, and A. Fornito. Consistencies and differences between centrality measures across distinct classes of networks. *PLoS ONE*, 14(7):0220061, 2019. doi:10.1371/journal.pone.0220061.

19 L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web, technical report. *Stanford Digital Library*, 1999.

20 K. Blackmond Laskey P.W. Holland and S. Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109–137, 1983. doi:10.1016/0378-8733(83)90021-7.

21 C. Orsini R. Aldecoa and D. Krioukov. Hyperbolic graph generator. *Computer Physics Communications*, 196:492–496, 2015. doi:10.1016/j.cpc.2015.05.028.

22 C. Sciarra, G. Chiarotti, F. Laio, and L. Ridolfi. A change of perspective in network centrality. *Scientific Reports*, 2018. doi:10.1038/s41598-018-33336-8.

23 P. Shakarian, A. Bhatnagar, A. Aleali, E. Shaabani, and R. Guo. *The Independent Cascade and Linear Threshold Models*, chapter 4, pages 35–48. Briefs in Computer Science. Springer, 2015. doi:10.1007/978-3-319-23105-1_4.

24 C. Spearman. The proof and measurement of association between two things. *AM. J. Psychol*, 15:88–103, 1904.

25 D. Wagner U. Brandes, M. Gaertler. Experiments on graph clustering algorithms. In *Algorithms - ESA 2003*, pages 568–579. Springer Berlin Heidelberg, 2003.

26 T.W. Valente, K. Corognes, C. Lakon, and E. Costenbader. How correlated are network centrality measures? *Connections*, 28(1):16–26, 2008.

27 S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications.* Cambridge University Press, 1994.

A Details on the correlation analysis (Section 5.2, Fig. 5)

We detail the correlation coefficients for the centrality measures summarized in Figure 5. In all the forthcoming tables, the Kendall coefficients (τ) are shown in the upper triangular part and the Spearman coefficients (ρ) in the lower triangular part. For the `Katz` measure, a $-$ indicates non-convergence.

Table 6 Correlation for the different centrality measures on the synthetic GRP networks of size 100K from the data set.

GRPa									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	0,5504	0,5569	0,5385	0,4199	0,8184	0,5200	0,3807	0,3361
FWLTR	0,6971	1	0,0987	0,4204	0,1049	0,5684	0,1379	0,0556	0,4694
BWLTR	0,7051	0,1335	1	0,4185	0,5962	0,5672	0,7634	0,6151	0,0882
Betweenness	0,7125	0,5655	0,5642	1	0,4899	0,6015	0,4401	0,4045	0,2871
Closeness	0,5782	0,1490	0,7608	0,6734	1	0,4775	0,7082	0,5674	0,1046
Degree	0,9228	0,7139	0,7135	0,7757	0,6442	1	0,5910	0,4314	0,3685
Katz	0,6943	0,1954	0,8995	0,6153	0,8868	0,7653	1	0,6170	0,1337
Pagerank	0,5286	0,0791	0,7757	0,5707	0,7573	0,5882	0,8040	1	0,0510
ICRt	0,4714	0,6248	0,1257	0,4174	0,1564	0,5106	0,1992	0,0764	1

GRPb									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	0,6088	0,6057	0,5473	0,4894	0,8402	-	0,4116	0,4075
FWLTR	0,7631	1	0,2149	0,4548	0,2271	0,6248	-	0,1240	0,5059
BWLTR	0,7609	0,2917	1	0,4505	0,6248	0,6176	-	0,6326	0,1957
Betweenness	0,7251	0,6105	0,6059	1	0,5109	0,5889	-	0,4138	0,3272
Closeness	0,6633	0,3213	0,7914	0,6988	1	0,5478	-	0,5411	0,2207
Degree	0,9398	0,7767	0,7707	0,7663	0,7266	1	-	0,4473	0,4453
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,5687	0,1774	0,7945	0,5826	0,7291	0,6107	-	1	0,1118
ICRt	0,5646	0,6691	0,2775	0,4717	0,3255	0,6078	-	0,1669	1

GRPc									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1,0000	0,5664	0,5656	0,7066	0,4633	0,9123	0,5180	0,4118	0,3212
FWLTR	0,7090	1,0000	0,0584	0,4872	0,0572	0,5639	0,0658	0,0371	0,4628
BWLTR	0,7081	0,0785	1,0000	0,4852	0,7039	0,5630	0,8367	0,6444	0,0423
Betweenness	0,8661	0,6397	0,6379	1,0000	0,5120	0,7322	0,5052	0,4609	0,3069
Closeness	0,6273	0,0812	0,8566	0,6954	1,0000	0,4805	0,8177	0,6689	0,0447
Degree	0,9706	0,7061	0,7051	0,8851	0,6464	1,0000	0,5410	0,4334	0,3310
Katz	0,6886	0,0934	0,9457	0,6885	0,9546	0,7125	1,0000	0,6833	0,0505
Pagerank	0,5647	0,0528	0,8022	0,6382	0,8515	0,5901	0,8621	1,0000	0,0272
ICRt	0,4487	0,6154	0,0600	0,4437	0,0669	0,4609	0,0755	0,0408	1,0000

GRPd									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	1	1	0,6530	0,6144	0,9391	-	0,7964	0,4652
FWLTR	1	1	1	0,6530	0,6144	0,9391	-	0,7964	0,4652
BWLTR	1	1	1	0,6530	0,6144	0,9391	-	0,7964	0,4652
Betweenness	0,8131	0,8131	0,8131	1	0,7164	0,6812	-	0,5996	0,3806
Closeness	0,7791	0,7791	0,7791	0,8904	1	0,6507	-	0,4734	0,4089
Degree	0,9767	0,9767	0,9767	0,8355	0,8112	1	-	0,8248	0,4816
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,9204	0,9204	0,9204	0,7949	0,6578	0,9356	-	1	0,3856
ICRt	0,6204	0,6204	0,6204	0,5423	0,5805	0,6368	-	0,5502	1

Table 7 Correlation for the different centrality measures on the synthetic LFR networks of size 100K from the data set.

LFRa									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	1	1	0,2514	0,5987	0,4142	-	0,2761	0,3573
FWLTR	1	1	1	0,2514	0,5987	0,4142	-	0,2761	0,3573
BWLTR	1	1	1	0,2514	0,5987	0,4142	-	0,2761	0,3573
Betweenness	0,3475	0,3475	0,3475	1	0,3881	0,4660	-	0,4483	0,2909
Closeness	0,7887	0,7887	0,7887	0,5414	1	0,4231	-	0,2739	0,3695
Degree	0,5541	0,5541	0,5541	0,6162	0,5650	1	-	0,8558	0,4711
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,4006	0,4006	0,4006	0,6211	0,3969	0,9513	-	1	0,3673
ICRt	0,5103	0,5103	0,5103	0,4200	0,5250	0,6245	-	0,5223	1

LFRb									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	1	1	0,3212	0,6461	0,6059	-	0,4276	0,5392
FWLTR	1	1	1	0,3212	0,6461	0,6059	-	0,4276	0,5392
BWLTR	1	1	1	0,3212	0,6461	0,6059	-	0,4276	0,5392
Betweenness	0,4371	0,4371	0,4371	1	0,4125	0,4852	-	0,4911	0,2999
Closeness	0,8310	0,8310	0,8310	0,5639	1	0,5194	-	0,3302	0,5080
Degree	0,7708	0,7708	0,7708	0,6360	0,6823	1	-	0,8205	0,5399
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,6007	0,6007	0,6007	0,6708	0,4781	0,9346	-	1	0,3934
ICRt	0,7313	0,7313	0,7313	0,4299	0,6970	0,7072	-	0,5576	1

LFRc									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	1	1	0,2362	0,5899	0,4007	-	0,2704	0,3440
FWLTR	1	1	1	0,2362	0,5899	0,4007	-	0,2704	0,3440
BWLTR	1	1	1	0,2362	0,5899	0,4007	-	0,2704	0,3440
Betweenness	0,3244	0,3244	0,3244	1	0,3893	0,4570	-	0,4397	0,2852
Closeness	0,7793	0,7793	0,7793	0,5465	1	0,4347	-	0,2962	0,3624
Degree	0,5389	0,5389	0,5389	0,6060	0,5790	1	-	0,8644	0,4682
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,3937	0,3937	0,3937	0,6115	0,4271	0,9571	-	1	0,3732
ICRt	0,4934	0,4934	0,4934	0,4126	0,5162	0,6212	-	0,5304	1

LFRd									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	1	1	0,6590	0,4128	0,9887	-	0,8734	0,2870
FWLTR	1	1	1	0,6590	0,4128	0,9887	-	0,8734	0,2870
BWLTR	1	1	1	0,6590	0,4128	0,9887	-	0,8734	0,2870
Betweenness	0,8076	0,8076	0,8076	1	0,6365	0,6616	-	0,5454	0,3300
Closeness	0,5506	0,5506	0,5506	0,8257	1	0,4094	-	0,2862	0,3536
Degree	0,9954	0,9954	0,9954	0,8098	0,5459	1	-	0,8875	0,2836
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,9616	0,9616	0,9616	0,7289	0,4135	0,9700	-	1	0,2139
ICRt	0,3940	0,3940	0,3940	0,4751	0,5029	0,3891	-	0,3120	1

Table 8 Correlation for the different centrality measures on the synthetic STO networks of size 100K from the data set.

STOa									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	0,7657	0,7628	0,5716	0,6591	0,9080	-	0,4816	0,5613
FWLTR	0,9074	1	0,5423	0,5310	0,5206	0,7807	-	0,3081	0,5954
BWLTR	0,9055	0,7070	1	0,5295	0,7270	0,7743	-	0,6175	0,4697
Betweenness	0,7554	0,7074	0,7057	1	0,5592	0,5797	-	0,4508	0,4155
Closeness	0,8374	0,6982	0,8849	0,7522	1	0,6940	-	0,4939	0,4592
Degree	0,9769	0,9171	0,9127	0,7624	0,8673	1	-	0,4874	0,5897
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,6572	0,4380	0,7894	0,6291	0,6795	0,6643	-	1	0,2558
ICRt	0,7332	0,7609	0,6326	0,5848	0,6333	0,7600	-	0,3751	1

STOb									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	0,5155	0,5248	0,5183	0,3539	0,7955	0,4616	0,3545	0,2871
FWLTR	0,6558	1	0,0324	0,3959	0,0163	0,5361	0,0325	0,0082	0,4472
BWLTR	0,6667	0,0437	1	0,3876	0,5706	0,5397	0,7924	0,6075	0,0219
Betweenness	0,6888	0,5341	0,5251	1	0,4637	0,6024	0,4124	0,3861	0,2573
Closeness	0,4937	0,0232	0,7332	0,6425	1	0,4131	0,6761	0,5815	0,0204
Degree	0,9048	0,6758	0,6805	0,7739	0,5645	1	0,5297	0,4191	0,3162
Katz	0,6261	0,0463	0,9181	0,5809	0,8604	0,6989	1	0,6660	0,0316
Pagerank	0,4942	0,0116	0,7682	0,5477	0,7719	0,5710	0,8475	1	0,0049
ICRt	0,4049	0,5974	0,0311	0,3756	0,0306	0,4403	0,0473	0,0074	1

STOc									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	0,7269	0,7263	0,7334	0,6384	0,9499	-	0,5323	0,4527
FWLTR	0,8733	1	0,4191	0,6181	0,4000	0,7270	-	0,2926	0,4975
BWLTR	0,8731	0,5586	1	0,6185	0,7524	0,7264	-	0,6999	0,3161
Betweenness	0,8962	0,7923	0,7926	1	0,6139	0,7385	-	0,5372	0,4079
Closeness	0,8163	0,5507	0,8997	0,8063	1	0,6472	-	0,6002	0,3081
Degree	0,9906	0,8734	0,8732	0,8995	0,8241	1	-	0,5406	0,4585
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,7153	0,4151	0,8604	0,7283	0,7927	0,7246	-	1	0,2152
ICRt	0,6103	0,6536	0,4390	0,5736	0,4408	0,6166	-	0,3174	1

STOd									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	0,7804	0,7778	0,6526	0,7271	0,9695	-	0,4735	0,6019
FWLTR	0,9169	1	0,5273	0,5893	0,5454	0,7764	-	0,2877	0,6185
BWLTR	0,9149	0,6908	1	0,5873	0,7723	0,7734	-	0,6139	0,4801
Betweenness	0,8342	0,7697	0,7673	1	0,5720	0,6519	-	0,4984	0,4514
Closeness	0,8909	0,7250	0,9162	0,7682	1	0,7355	-	0,4813	0,5051
Degree	0,9955	0,9141	0,9119	0,8331	0,8968	1	-	0,4755	0,6071
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,6491	0,4102	0,7859	0,6831	0,6661	0,6511	-	1	0,2500
ICRt	0,7751	0,7846	0,6471	0,6310	0,6864	0,7797	-	0,3674	1

Table 9 Correlation for the different centrality measures on the synthetic HYP networks of size 100K from the data set.

HYPa									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	1	1	0,3516	0,8978	0,5677	-	0,2327	0,4664
FWLTR	1	1	1	0,3516	0,8978	0,5677	-	0,2327	0,4664
BWLTR	1	1	1	0,3516	0,8978	0,5677	-	0,2327	0,4664
Betweenness	0,4394	0,4394	0,4394	1	0,3377	0,5672	-	0,5245	0,3816
Closeness	0,9806	0,9806	0,9806	0,4222	1	0,5300	-	0,1856	0,4476
Degree	0,7015	0,7015	0,7015	0,6464	0,6621	1	-	0,6918	0,5833
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,3285	0,3285	0,3285	0,6465	0,2690	0,7802	-	1	0,3402
ICRt	0,6298	0,6298	0,6298	0,4741	0,6093	0,7160	-	0,4651	1

HYPb									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	1	1	0,3957	0,9618	0,5102	-	0,3386	0,4049
FWLTR	1	1	1	0,3957	0,9618	0,5102	-	0,3386	0,4049
BWLTR	1	1	1	0,3957	0,9618	0,5102	-	0,3386	0,4049
Betweenness	0,4999	0,4999	0,4999	1	0,3925	0,7631	-	0,6952	0,4380
Closeness	0,9963	0,9963	0,9963	0,5004	1	0,5113	-	0,3290	0,4099
Degree	0,6424	0,6424	0,6424	0,8707	0,6483	1	-	0,8014	0,5697
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,4734	0,4734	0,4734	0,8456	0,4689	0,8943	-	1	0,4135
ICRt	0,5575	0,5575	0,5575	0,5696	0,5671	0,7036	-	0,5675	1

HYPc									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	1	1	0,7696	0,5854	0,8678	-	0,7839	0,4477
FWLTR	1	1	1	0,7696	0,5854	0,8678	-	0,7839	0,4477
BWLTR	1	1	1	0,7696	0,5854	0,8678	-	0,7839	0,4477
Betweenness	0,9005	0,9005	0,9005	1	0,5989	0,7588	-	0,7111	0,4204
Closeness	0,7523	0,7523	0,7523	0,7789	1	0,5138	-	0,3994	0,4409
Degree	0,9358	0,9358	0,9358	0,8895	0,6668	1	-	0,8953	0,4276
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,9043	0,9043	0,9043	0,8817	0,5648	0,9704	-	1	0,3605
ICRt	0,5988	0,5988	0,5988	0,5853	0,6128	0,5692	-	0,5090	1

HYPd									
p \ t	LTR	FWLTR	BWLTR	Betweenness	Closeness	Degree	Katz	Pagerank	ICRt
LTR	1	1	1	0,4444	0,3961	0,7839	-	0,5402	0,4800
FWLTR	1	1	1	0,4444	0,3961	0,7839	-	0,5402	0,4800
BWLTR	1	1	1	0,4444	0,3961	0,7839	-	0,5402	0,4800
Betweenness	0,5921	0,5921	0,5921	1	0,1939	0,5260	-	0,5842	0,2417
Closeness	0,5447	0,5447	0,5447	0,2787	1	0,2962	-	0,1065	0,4101
Degree	0,8890	0,8890	0,8890	0,6754	0,4082	1	-	0,7186	0,4474
Katz	-	-	-	-	-	-	-	-	-
Pagerank	0,6992	0,6992	0,6992	0,7652	0,1587	0,8549	-	1	0,2560
ICRt	0,6415	0,6415	0,6415	0,3468	0,5713	0,5931	-	0,3673	1