HAL open science

Alpha current flow betweenness centrality

Konstantin Avrachenkov, Nelly Litvak, Vasily Medyanikov, Marina Sokol

To cite this version:

Konstantin Avrachenkov, Nelly Litvak, Vasily Medyanikov, Marina Sokol. Alpha current flow betweenness centrality. WAW - 10th International Workshop on Algorithms and Models for the Web Graph, Dec 2013, Cambridge, Massachusetts, United States. pp.106-117, 10.1007/978-3-319-03536-9_9 . hal-00926438

HAL Id: hal-00926438

https://inria.hal.science/hal-00926438v1

Submitted on 9 Jan 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Alpha current flow betweenness centrality*

Konstantin Avrachenkov[1], Nelly Litvak[2], Vasily Medyanikov[3], Marina Sokol[1]

[1] Inria Sophia Antipolis, 2004 Route des Lucioles, Sophia-Antipolis, France
[2] University of Twente, P.O.Box 217, 7500AE, Enschede, The Netherlands
[3] St. Petersburg State University, 7-9, Universitetskaya nab., St. Petersburg, Russia

Abstract. A class of centrality measures called betweenness centralities reflects degree of participation of edges or nodes in communication between different parts of the network. The original shortest-path betweenness centrality is based on counting shortest paths which go through a node or an edge. One of shortcomings of the shortest-path betweenness centrality is that it ignores the paths that might be one or two hops longer than the shortest paths, while the edges on such paths can be important for communication processes in the network. To rectify this shortcoming a current flow betweenness centrality has been proposed. Similarly to the shortest-path betweenness, it has prohibitive complexity for large size networks. In the present work we propose two regularizations of the current flow betweenness centrality, α-current flow betweenness and truncated α-current flow betweenness, which can be computed fast and correlate well with the original current flow betweenness.

1 Introduction

A class of centrality measures called betweenness centralities reflects degree of participation of edges or nodes in communication between different parts of the network. The first notion of betweenness centrality was introduced by Freeman [8]. Let $s, t \in V$ be a pair of nodes in an undirected graph $G = (V, E)$. (In the present work we restrict our consideration to undirected graphs.) We denote $|V| = n$, $|E| = m$, and let d_v be the degree of node v. Let $\sigma_{s,t}$ be the number of shortest paths connecting nodes s and t and denote by $\sigma_{s,t}(e)$ the number of shortest paths connecting nodes s and t passing through edge e. Then betweenness centrality of edge e is calculated as follows:

$$C_{\mathrm{B}}(e) = \frac{1}{n(n-1)} \sum_{s,t \in V} \frac{\sigma_{s,t}(e)}{\sigma_{s,t}}. \tag{1}$$

Computational complexity of the best known algorithm for computing the betweenness in (1) is $\mathcal{O}(mn)$ [4]. This limits its applicability for large graphs.

* This research is partially funded by Inria Alcatel-Lucent Joint Lab, by the European Commission within the framework of the CONGAS project FP7-ICT-2011-8-317672, see www.congas-project.eu, and by the EU-FET Open grant NADINE (288956).

One of shortcomings of the betweenness centrality in (1) is that it takes into accounts only the shortest paths, ignoring the paths that might be one or two steps longer, while the edges on such paths can be important for communication processes in the network. In order to take such paths into account, Newman [12] and Brandes and Fleischer [5] introduced the current flow betweenness centrality (CF-betweenness). In [12, 5] the graph is regarded as an electrical network with edges being unit resistances. The CF-betweenness of an edge is the amount of current that flows through it, averaged over all source-destination pairs, when one unit of current is induced at the source, and the destination (sink) is connected to the ground. This exploits the well known relation between electrical networks and reversible Markov chains, see e.g., [1, 7].

The computational difficulty of Betweenness and the CF-betweenness is that the computations must be done over the set of all source-destination pairs. The best previously known computational complexity for the CF-betweenness is $\mathcal{O}(I(n-1) + mn \log n)$ where $I(n-1)$ is the complexity of the inversion of matrix of dimension $n-1$.

In the present work we introduce new betweenness centrality measures: α-current flow betweenness (α-CF betweenness) and its truncated version. The main purpose of these new measures is to bring down the high cost of the CF-betweenness computation. Our proposed measures are very close in performance to the CF-betweenness, but they are comparable to the PageRank algorithm [6] in their modest computational complexity. Our goal is to provide and analyze efficient algorithms for computing α-CF betweenness and truncated α-CF betweenness, and to compare the α-CF betweenness to other centrality measures.

2 Alpha current flow betweenness

We view the graph G as an electrical network where each edge has resistance α^{-1}, and each node v is connected to ground node $n+1$ by an edge with resistance $(1-\alpha)^{-1}d_v^{-1}$. This is in the spirit of PageRank. Indeed, the current (probability flow) is inversely proportional to the resistance. Thus, the fraction α of the current from node v flows to the network, while fraction $(1-\alpha)$ of the current is directed to the sink. Since the graph is undirected, we use a convention that (v, w) and (w, v) represent the same arc in E, but depending on the chosen direction the current along this arc is considered to be positive or negative.

Assume that a unit of current is supplied to a source node $s \in V$, and there is a destination node $t \in V$ connected to the ground. Let $\varphi_v^{(s,t)}$ denote the absolute potential of node $v \in V$, if s is the source and t is the destination. Assume without loss of generality that $s = 1$ and $t = n$ ($\varphi_n^{(1,n)} = \varphi_{n+1}^{(1,n)} = 0$, i.e., we set the ground potential to zero). The vector of absolute potentials of the other nodes $\varphi^{(1,n)} = [\varphi_1^{(1,n)}, ..., \varphi_{n-1}^{(1,n)}]^T$ is a solution of the following system of equations (Kirchhoff's current law):

$$[\tilde{D} - \alpha\tilde{A}]\varphi^{(1,n)} = \tilde{b}, \tag{2}$$

where $\tilde{D}$ and $\tilde{A}$ are the degree and adjacency matrices of the graph without node n and $\tilde{b} = [1, 0, ..., 0]^T$, see e.g., [5].

In the following theorem we demonstrate that we do not need to solve a separate linear system for each source-destination pair, it suffies to invert the coefficient matrix $[D - \alpha A]$.

Theorem 1 *The voltage drop along the edge (v, w) is given by*

$$\varphi_v^{(s,t)} - \varphi_w^{(s,t)} = (c_{s,v} - c_{s,w}) + \frac{c_{s,t}}{c_{t,t}}(c_{t,w} - c_{t,v}), \tag{3}$$

where $(c_{v,w})_{v,w \in V}$, are the elements of the matrix $C = [D - \alpha A]^{-1}$.

Proof: Assume again without loss of generality that $s = 1$ and $t = n$. The matrix $[D - \alpha A]$ can be written in the following block structure

$$D - \alpha A = \begin{bmatrix} \tilde{D} - \alpha\tilde{A} & -\alpha\tilde{a} \\ -\alpha\tilde{a}^T & d_n \end{bmatrix}, \quad \text{with} \quad \tilde{a} = [a_{1,n}, a_{2,n}, \ldots, a_{n-1,n}]^T.$$

Then, divide accordingly the elements of the inverse matrix

$$C = [D - \alpha A]^{-1} = \begin{bmatrix} \tilde{C} & \tilde{c} \\ \tilde{c}^T & c_{n,n} \end{bmatrix}.$$

Writing the relation $[D - \alpha A]C = I$ in the block form yields

$$[\tilde{D} - \alpha\tilde{A}]\tilde{C} - \alpha\tilde{a}\tilde{c}^T = I, \tag{4}$$

$$[\tilde{D} - \alpha\tilde{A}]\tilde{c} - \alpha\tilde{a}\tilde{c}_{n,n} = 0. \tag{5}$$

Premultiplying equation (4) by $[\tilde{D} - \alpha\tilde{A}]^{-1}$, we obtain

$$[\tilde{D} - \alpha\tilde{A}]^{-1} = \tilde{C} - \alpha[\tilde{D} - \alpha\tilde{A}]^{-1}\tilde{a}\tilde{c}^T. \tag{6}$$

And premultiplying (5) by $[\tilde{D} - \alpha\tilde{A}]^{-1}$, we obtain

$$\alpha[\tilde{D} - \alpha\tilde{A}]^{-1}\tilde{a} = \frac{1}{c_{n,n}}\tilde{c}. \tag{7}$$

Combining both equations (6) and (7) gives

$$[\tilde{D} - \alpha\tilde{A}]^{-1} = \tilde{C} - \frac{1}{c_{n,n}}\tilde{c}\tilde{c}^T,$$

and hence $\varphi^{(1,n)} = [\tilde{D} - \alpha\tilde{A}]^{-1}\tilde{b} = \tilde{C}_{\cdot,1} - \frac{c_{1,n}}{c_{n,n}}\tilde{c}$. Thus, we can write

$$\varphi_v^{(1,n)} - \varphi_w^{(1,n)} = (c_{v,1} - c_{w,1}) + \frac{c_{1,n}}{c_{n,n}}(c_{w,n} - c_{v,n})$$

The above expression is symmetric and can be rewritten for any source-target pair (s,t). That is,

$$\varphi_v^{(s,t)} - \varphi_w^{(s,t)} = (c_{v,s} - c_{w,s}) + \frac{c_{s,t}}{c_{t,t}}(c_{w,t} - c_{v,t}).$$

Furthermore, since matrix C is symmetric for undirected graphs, we can rewrite the above equation as

$$\varphi_v^{(s,t)} - \varphi_w^{(s,t)} = (c_{s,v} - c_{s,w}) + \frac{c_{s,t}}{c_{t,t}}(c_{t,w} - c_{t,v}),$$

which completes the proof. □

The potentials $\varphi_v^{(s,t)}$, $v,s,t \in V$, have a clear probabilistic interpretation. Take again $s=1$ and $t=n$. Then from (2) we readily obtain

$$\varphi_v^{(1,n)} = \mathbf{e}_v^T[\tilde{D} - \alpha\tilde{A}]^{-1}\tilde{b} = \mathbf{e}_v^T[I - \alpha\tilde{P}]^{-1}\tilde{D}^{-1}\tilde{b}, \tag{8}$$

where $\mathbf{e}_v$ is a v-th standard basis column vector, and $\tilde{P}$ is the transition probability matrix for a simple random walk on G with absorption in n. Compare this to the well-known expression for the Personalized PageRank vector $\pi(v) = (\pi_1(v), \ldots, \pi_n(v))$ with teleportation preference concentrated in v and damping factor α: $\pi(v) = (1-\alpha)\mathbf{e}_v^T[I - \alpha P]^{-1}$. Note that the vector $\tilde{\pi}(v) = (1-\alpha)\mathbf{e}_v^T[I - \alpha\tilde{P}]^{-1}$ is very similar to $\pi(v)$, except it nullifies the contribution of node n. Now, recall that $\tilde{b} = (1,0,\ldots,0)^T$ to obtain

$$\varphi_v^{(1,n)} = (1-\alpha)^{-1}\tilde{\pi}_1(v)d_1^{-1}.$$

Furthermore, let $\mathbf{1}$ be a column vector of ones. Recall that the PageRank vector with uniform teleportation can be written as $\pi = \frac{1-\alpha}{n}\mathbf{1}^T[I - \alpha P]^{-1}$, and define a similar vector $\tilde{\pi} = \frac{1-\alpha}{n}\mathbf{1}^T[I - \alpha\tilde{P}]^{-1}$. Then

$$\sum_{v \in V} \varphi_v^{(1,n)} = n(1-\alpha)^{-1}\tilde{\pi}_1 d_1^{-1}.$$

It is well-known (see e.g., [9] and references therein) and is also confirmed by our experiments that the PageRank of a node in an undirected graph is strongly correlated to the degree of the node. Thus, with any choice of the source, the sum of the potentials is of similar magnitude, except for the cases when the destination node has a large contribution into the PageRank mass of the source.

Finally, we note that the source node has the highest potential, and from [1, Chapter 3, Section 3] we find

$$\begin{aligned}\varphi_1^{(1,n)} &= [P(\text{random walk returns to node 1 before absorption})]^{-1} \\ &= E(\#\text{ returns to node 1 before absorption}).\end{aligned}$$

Now we are ready to define α-CF betweenness. The current $I_e^{(s,t)}$ through edge $e=(v,w)$ is equal to $\alpha(\varphi_v^{(s,t)} - \varphi_w^{(s,t)})$. Let

$$x_e^{(s,t)} = |\varphi_v^{(s,t)} - \varphi_w^{(s,t)}|, \quad (v,w) \in E,$$

be the difference of potentials, that determines the absolute value of the current on the edge. The α-CF betweenness of edge e is defined by

$$x_e^\alpha = \frac{1}{n(n-1)} \sum_{s,t \in V, s \neq t} x_e^{(s,t)}, \quad e \in E. \tag{9}$$

Further, for each node $v \in V$ its α-CF betweenness is defined as the sum of the α-CF betweenness scores of its adjacent edges:

$$\alpha\text{-CF betweenness}(v) = \sum_{(v,w) \in E} x_{(v,w)}^\alpha, \quad v \in V. \tag{10}$$

With this definition, the node is central if a relatively large amount of current flows from this node to the network. This is in accordance to the original CF-betweenness of [12, 5], except we introduced the additional sink ground node $n+1$. This mitigates the computational issues because the original CF-betweenness requires the inversion of the ill-conditioned matrix $[\tilde{D} - \tilde{A}]$, while for computing α-CF betweenness we need to invert the matrix $[D - \alpha A]$, which is a well posed problem, and has many efficient solutions (e.g., power iteration and Monte Carlo methods). In fact, as we shall show below, we need to obtain just a few rows of the inverse matrix $[D - \alpha A]^{-1}$. In the rest of the paper we will discuss the computation and the properties of the α-CF betweenness.

3 Computation of α-CF betweenness

Due to the presence of the auxiliary node $n+1$, the value of $x_e^{(s,t)}$ on the right-hand side of (9) can be computed efficiently with high precision for any source-destination pair. However, the summation over all $n(n-1)$ pairs is a problem of prohibitive computational complexity even for graphs of modest size. The solution is to perform the computations for sufficiently many source-destination pairs. Since all source-destination pairs contribute equally in (9) we choose to sample them uniformly at random. This results in the next algorithm for computing the α-CF betweenness.

Algorithm 1.

1. Select a set of pairs of nodes $(s_i, t_i), i = 1, ..., N$, uniformly at random;
2. For each s_i and t_i, $i = 1, ..., N$ compute the rows $c_{s_i,\cdot}$, $c_{t_i,\cdot}$ (this can be done either by power iteration or by Monte Carlo algorithm);
3. For each edge $e = (v, w)$ and each pair (s_i, t_i), use (3) to compute

$$x_e^{(s_i,t_i)} = |\varphi_v^{(s_i,t_i)} - \varphi_w^{(s_i,t_i)}|.$$

4. Average over source-destination pairs

$$\bar{x}_e^\alpha = \frac{1}{N} \sum_{i=1}^{N} x_e^{(s_i,t_i)}.$$

Since we chose the pairs (s_i, t_i) uniformly at random then for every edge e, $\bar{x}_e^\alpha$ is just a sample average where all values are between zero and one. Then using the standard approach for the analysis of the series of independent random variables we have the following result.

Theorem 2 *Algorithm 1 approximates the alpha current flow betweenness in $O(m \log(n)\varepsilon^{-2} \log(\varepsilon)/\log(\alpha))$ time to within an absolute error of ε with arbitrarily high fixed probability.*

Proof: In addition to the proof of Theorem 3 in [5] we just need to note that we can compute Personalized PageRank with precision ε in $\mathrm{O}(\log(\varepsilon)/\log(\alpha))$ power iterations. □

4 Truncated α-CF betweenness

In the experiments we noticed that the values $x_e^{(s,t)}$ have a high variance, which results in poor precision when evaluating x_e^α. A closer analysis revealed that the edges adjacent to the source s receive large values of $x_e^{(s,t)}$, especially when $e = (v, s)$, where v has degree 1, so (v, s) is its only edge, and s has a large degree. This can be explained using the random walk interpretation. Consider a PageRank-type random walk on G. At each node, with probability α, the random walk traverses a randomly chosen edge of this node, and with probability $1-\alpha$ it jumps to the sink, node $n+1$. Denote by T_B the number of steps of the random walk needed to hit set B. It follows from Proposition 10 of [1, Chapter 3] that $\varphi_v^{(s,t)}/\varphi_s^{(s,t)} = P_v(T_{\{s\}} < T_{\{t,n+1\}})$, where $P_v(\cdot)$ is a conditional probability given that the random walk starts at v. Hence, if s is the only neighbor of v then $\varphi_v^{(s,t)}/\varphi_s^{(s,t)} = \alpha$, the probability of no absorption before reaching s. Thus, $|\varphi_s^{(s,t)} - \varphi_v^{(s,t)}| = (1-\alpha)\varphi_s^{(s,t)}$, which can be large if e.g. $\alpha = 0.8$ because $\varphi_s^{(s,t)}$ is the largest potential in the network. In contrast, the original CF-betweenness corresponds to $\alpha = 1$, implying that the current in (v, s) is zero.

This prompts us to introduce the truncated version of α-CF betweenness where for each edge (v, w) we only take into account the scores $x_{(v,w)}^{(s,t)}$ if $v, w \neq s$. In Fig. 1 we present log-linear plots of the empirical complementary distribution function of $x_{(v,w)}^{(s,t)}$ over all pairs (s, t) (solid line), and its truncated version (dashed line). The plots are given for two edges in the Dolphin social network described in Section 5 below. Nodes 1 and 36 are central in the network, so the high α-CF betweenness of (1,36) is expected. Node 60 has degree 1, so edge (32,60) gains an unwanted high betweenness in the non-truncated version.

Since the truncated α-CF betweenness gives lower scores to the edges connected to nodes of degree 1, one can expect that it has a higher correlation with CF-betweenness, especially for middle-range values of α. This is confirmed below in Fig. 2. Moreover, the truncated version removes outliers, and does not have large spread in values, thus standard statistical procedures, based on the Central Limit Theorem can be applied. Also, because of the smaller variance, Algorithm 1 achieves a desired precision with a smaller sample of source-destination pairs.

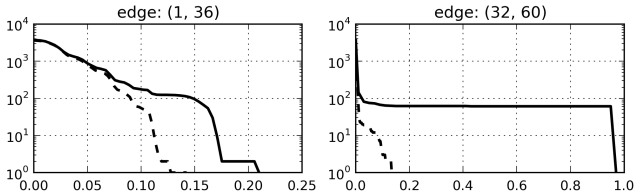

Fig. 1. The number of pairs s, t with $x_{(v,w)}^{(s,t)} > x$ over all pairs (s, t) (solid line) and only pairs with $v, w \neq s$ (dashed line).

5 Datasets

We consider the four graphs described below.

Dolphin social network. This small graph represents a social network of frequent associations between 62 dolphins in a community living off Doubtful Sound, New Zealand [11].

Graph of VKontakte social network. We have collected data from a popular Russian social network VKontakte. We were considering subgraph representing one of the connected components of people who stated that they were studying at Applied Mathematics - Control Processes Faculty at the St. Petersburg State University in different years. We ran the breadth-first search (BFS) algorithm starting at one specific node of the network and then anonymized the obtained users' data leaving only information about connections between people. Collected network consists of 2092 individuals out of total 8859 denoted the specified faculty in the Education field.

Watts-Strogatz model. As an artificial example, we used a random graph generated by the Watts-Strogatz model. We have chosen this model as it combines high clustering and short average path length, thus different centrality measures give very different results on this graph. For other random models considered (Erdös-Rényi and Barabási-Albert) all measures are highly correlated and behave very similarly to each other.

Enron graph. Enron email communication network is a well known test dataset. It covers all the email communication within a dataset of around half million emails between Enron's employees. The node are e-mail addresses, and an edge appears if an e-mail message was sent from one e-mail to another. Although this graph is small compared to, say, web or Twitter samples, it is already prohibitively large for computing the CF-betweenness in its original form.

6 Numerical results for α-CF betweenness

To begin with, we compare the two versions of α-CF betweenness (truncated and without truncation) to the CF-betweenness scores defined as in [12, 5]. Fig. 2 presents the results for the three smaller graphs, in which the latter measure

	$\|V\|$	$\|E\|$	$\langle deg(v) \rangle$	$diam(G)$	$C_{\text{clustering}}$	$\langle d(u,v) \rangle$
Dolphin social network	62	159	5.13	8	0.259	3.357
VKontakte AMCP social graph	2092	14816	14.16	14	0.338	4.598
Watts-Strogatz	1000	6000	12.00	6	0.422	3.713
$(n = 1000, k = 12, p = 0.150)$						
Enron	36692	183831	10.02	11	0.4970	≈ 4.8

Table 1. Datasets characteristics

could be computed. As a correlation measure we use the Kendall tau rank correlation. We observe that the truncated version is better correlated with the

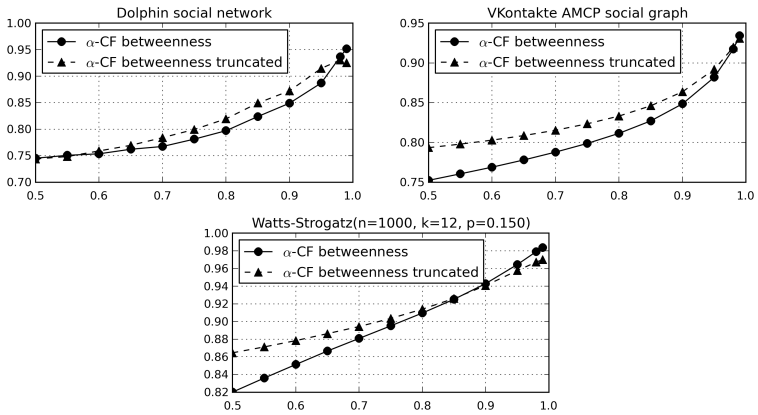

Fig. 2. Correlations between α-CF betweenness and truncated α-CF betweenness with CF-betweenness as a function of α.

CF-betweenness when α is not very close to one. As explained above, this is because the high probability of absorption in the auxiliary node $n + 1$ results in a relatively high current in the edges connected to the source, which is not necessarily the case if absorption is only possible in the destination node.

Next, we demonstrate that we can compute α-CF betweenness in the Enron graph, where the computation of CF-betweenness is infeasible (at least, with our means). We have evaluated α-CF betweenness, non-truncated and truncated, with $\alpha = 0.98$. We have run Algorithm 1 using $N = 20 \cdot 10^6$ source-destination pairs. In Fig. 3 we plot the complementary distribution function in log-linear scale, of the score $x_e^{0.98}$ across the edges.

Note that distribution over edges (the left plot in Fig. 3) does not have a large spread of values, except one outlier edge that connects two most important hubs. Since the weights of the edges are comparable, it is to be expected that in this graph the nodes of large degrees are also the ones with highest betweenness.

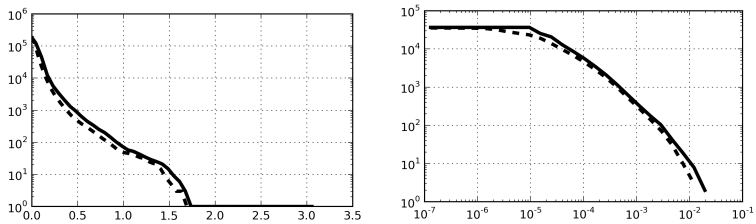

Fig. 3. Distribution of α-CF betweenness scores in the Enron graph, truncated (dashed line) and not truncated (solid line). Left: $10^4 \cdot x_e^{0.98}$ for edges $e \in E$. Right: α-CF betweenness (v) for $v \in V$. On the x-axis are the values of α-CF betweenness, on the y-axis the number of edges/nodes with the score larger than x.

Indeed, the Kendall's tau correlation between α-CF betweenness and degree of the nodes turns out to be 0.808, which is higher than in the three smaller graphs. The reason can be either the graph size or its structure. In future research we will investigate how the CF-betweenness score, e.g. its maximum value across the edges, scales with the graph size in graphs with power law degrees.

We further present correlations between our proposed measures and other measures of betweenness. These are computed on smaller graphs where we could obtain exact values of all presented measures, see Tables 2–4. For completeness, we have also included PageRank (PR) computed with $\alpha = 0.85$ and a distance-base centrality measure – Closeness centrality:

$$C_{\mathrm{C}}(v) = \frac{n-1}{\sum_{w \in V, w \neq v} d(v,w)},$$

where $d(v,w)$ is the graph distance between v and w. Betweenness (Between.) is computed as in (1).

	Degree	PR	Closeness	Between.	CF	αCF(0.8)	αCF-tr(0.8)	αCF(0.98)
Degree	1.000	0.930	0.548	0.665	0.737	0.864	0.855	0.769
PageRank	0.930	1.000	0.458	0.658	0.733	0.872	0.827	0.757
Closeness	0.548	0.458	1.000	0.578	0.575	0.515	0.573	0.591
Between.	0.665	0.658	0.578	1.000	0.829	0.749	0.759	0.828
CF	0.737	0.733	0.575	0.829	1.000	0.798	0.820	0.939
αCF(0.8)	0.864	0.872	0.515	0.749	0.798	1.000	0.925	0.838
αCF-tr(0.8)	0.855	0.827	0.573	0.759	0.820	0.925	1.000	0.876
αCF(0.98)	0.769	0.757	0.591	0.828	0.939	0.838	0.876	1.000

Table 2. Kendall tau for centrality measures in Dolphin social network.

Note that α-CF betweenness is strongly correlated with CF-betweenness. The Closeness Centrality does not agree well with the CF-betweenness, even the

	Degree	PR	Closeness	Between.	CF	αCF(0.8)	αCF-tr(0.8)	αCF(0.98)
Degree	1.000	0.655	0.679	0.521	0.545	0.659	0.668	0.599
PageRank	0.655	1.000	0.375	0.662	0.717	0.833	0.811	0.766
Closeness	0.679	0.375	1.000	0.382	0.356	0.424	0.445	0.395
Between.	0.521	0.662	0.382	1.000	0.761	0.760	0.749	0.778
CF	0.545	0.717	0.356	0.761	1.000	0.812	0.833	0.917
αCF(0.8)	0.659	0.833	0.424	0.760	0.812	1.000	0.938	0.878
αCF-tr(0.8)	0.668	0.811	0.445	0.749	0.833	0.938	1.000	0.903
αCF(0.98)	0.599	0.766	0.395	0.778	0.917	0.878	0.903	1.000

Table 3. Kendall tau for centrality measures in the social graph VKontakte AMCP.

	Degree	PR	Closeness	Between.	CF	αCF(0.8)	αCF-tr(0.8)	αCF(0.98)
Degree	1.000	0.891	0.462	0.526	0.610	0.643	0.581	0.612
PageRank	0.891	1.000	0.415	0.485	0.565	0.610	0.546	0.567
Closeness	0.462	0.415	1.000	0.655	0.613	0.647	0.666	0.628
Between.	0.526	0.485	0.655	1.000	0.853	0.819	0.852	0.857
CF	0.610	0.565	0.613	0.853	1.000	0.910	0.914	0.979
αCF(0.8)	0.643	0.610	0.647	0.819	0.910	1.000	0.935	0.923
αCF-tr(0.8)	0.581	0.546	0.666	0.852	0.914	0.935	1.000	0.930
αCF(0.98)	0.612	0.567	0.628	0.857	0.979	0.923	0.930	1.000

Table 4. Kendall tau for centrality measures in the Watts-Strogatz graph (n=1000, k=12, p=0.150).

PageRank and the degrees have a higher correlations with the CF-betweenness in real graphs. Recent paper [2] suggests more measures based on distance, and efficient computation methods for such measures is presented in [3]. In future it will be interesting to compare these new measures to α-CF betweenness.

7 Centrality measures and network vulnerability

We now consider how well the CF-betweenness and α-CF betweenness can indicate the nodes responsible for maintaining connectivity of a network. We follow the methodology of [10]. As measures of connectivity we choose the average inverse distance

$$< d^{-1} >= \frac{1}{n(n-1)} \sum_{u,v \in V, u \neq v} \frac{1}{d(u,v)}$$

and the size of the largest connected component. In the experiment, we remove the top nodes one by one, according to different betweenness measures, and observe how the connectivity of the network changes. In Fig. 4 the results are presented for the inversed average distance.

The results for the social graph VKontakte are especially interesting, because this network turns out to be less vulnerable to the removal of nodes with large degree than nodes with large betweenness and its modifications (CF-betweenness,

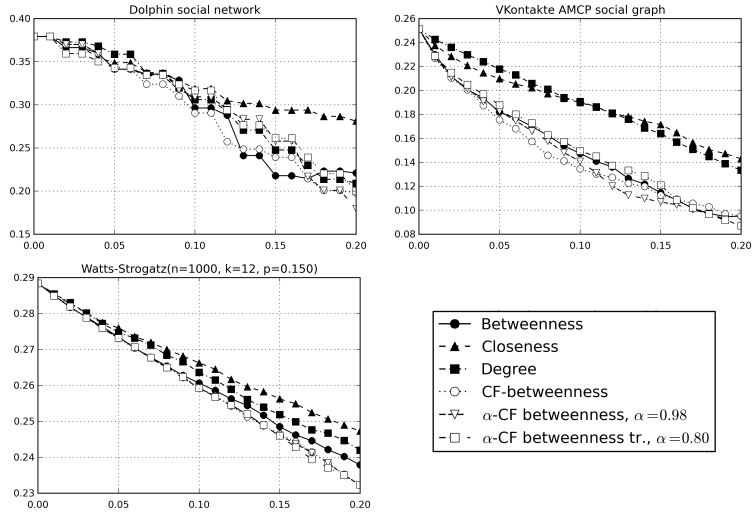

Fig. 4. Inverse average distance as a function of the fraction of removed top-nodes according to different betweenness centrality measures.

α-CF betweenness, and truncated α-CF betweenness). On the small Dolphin social network there is no much difference in vulnerability with respect to different centrality measures. Finally, on the artificial Watts-Strogatz graph the CF-betweenness and our proposed two versions of α-CF betweenness find the nodes that are most essential for the network connectivity.

In Fig. 5 we plot the size of the largest connected components against the fraction of removed top-nodes. We do not present the plot for the Watts-Strogatz graph because it remains entirely connected, so the size of its largest connected component equals to the number of remaining nodes irrespectively of which nodes are removed first. For the two real graphs, the CF-betweenness is most efficient in reducing the size of the giant component. On the Dolphin graph, α-CF betweenness performs closely to CF-betweenness, except the interval when 13-18% of nodes are removed. On the graph VKontakte, α-CF betweenness and its truncated version perform comparably to the CF-betweenness. Again, on this graph, degree and Closeness centrality fail to reveal the nodes responsible for good network connectivity. The α-CF betweenness with $\alpha = 0.98$ appears to be a better indicator for vulnerability than the truncated α-CF betweenness with $\alpha = 0.8$. The latter however also gives good results, and can be computed easier on large graphs due to the faster convergence of the power iteration algorithm.

We conclude that both α-CF betweenness and truncated α-CF betweenness provide an adequate measure for the role of a node in network's connectivity. Furthermore, their computational costs are lower than for known measures of betweenness, and the computations can be done in parallel easily. Thus, α-CF betweenness can be applied in large graphs, for which computation of other measures of betweenness is merely infeasible.

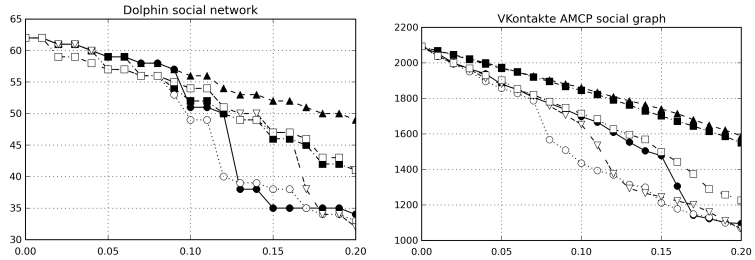

Fig. 5. The size of the largest connected component as a function of the fraction of removed top-nodes according to different betweenness centrality measures. (same legend as in Fig. 4)

References

1. D. Aldous and J. Fill. Reversible Markov chains and random walks on graphs. 1999.
2. P. Boldi and S. Vigna. Axioms for centrality. *arXiv:1308.2140*.
3. P. Boldi and S. Vigna. In-core computation of geometric centralities with hyperball: A hundred billion nodes and beyond. *arXiv:1308.2144*.
4. U. Brandes. A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(1994):163–177, 2001.
5. U. Brandes and D. Fleischer. Centrality measures based on current flow. In *Proceedings of the 22nd annual conference on Theoretical Aspects of Computer Science*, pages 533–544, 2005.
6. S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and {ISDN} Systems*, 30(17):107–117, 1998.
7. P.G. Doyle and J.L. Snell. *Random walks and electric networks*. Mathematical Association of America, 1984.
8. L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 1977.
9. V. Grolmusz. A note on the pagerank of undirected graphs. *arXiv preprint arXiv:1205.1960*, 2012.
10. P. Holme, B.J. Kim, C.N. Yoon, and S.K. Han. Attack vulnerability of complex networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 65(5 Pt 2):056109, May 2002.
11. D. Lusseau, K. Schneider, O.J. Boisseau, P. Haase, E. Slooten, and S.M. Dawson. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behavioral Ecology and Sociobiology*, 54(4):396–405, September 2003.
12. M.E.J. Newman. A measure of betweenness centrality based on random walks. *Social networks*, pages 1–15, 2005.