On the Local Approximations of Node Centrality in Internet Router-level Topologies

Technical Report

Panagiotis Pantazopoulos Merkourios Karaliopoulos and Ioannis Stavrakakis
Department of Informatics and Telecommunications, University of Athens, Ilissia, 157 84 Athens, Greece
Email: {ppantaz, mkaralio, ioannis}@di.uoa.gr

***Abstract*—In many networks with distributed operation and self-organization features, acquiring their global topological information is impractical, if feasible at all. Internet protocols drawing on node centrality indices may instead approximate them with their egocentric counterparts, computed out over the nodes' *ego-networks*. Surprisingly, however, in router-level topologies the approximative power of localized *ego-centered measurements* has not been systematically evaluated. More importantly, it is unclear how to practically interpret any positive correlation found between the two centrality metric variants.**

The paper addresses both issues using different datasets of ISP network topologies. We first assess how well the *egocentric* metrics approximate the original *sociocentric* ones, determined under perfect network-wide information. To this end we use two measures: their rank-correlation and the overlap in the top-k node lists the two centrality metrics induce. Overall, the rank-correlation is high, in the order of 0.8-0.9, and, intuitively, becomes higher as we relax the ego-network definition to include the ego's r-hop neighborhood. On the other hand, the top-k node overlap is low, suggesting that the high rank-correlation is mainly due to nodes of lower rank. We then let the node centrality metrics drive elementary network operations, such as local search strategies. Our results suggest that, even under high rank-correlation, the locally-determined metrics can hardly be effective aliases for the global ones. The implication for protocol designers is that rank-correlation is a poor indicator for the approximability of centrality metrics.

I. Introduction

Computer networks constitute a paradigm of a complex system, where complexity may arise due to the topological structure and node diversity or dynamical evolution processes. Understanding the behavior of networked interacting elements based on the measured statistical or local structural properties of given nodes, is the main focus of the highly interdisciplinary emerging framework of *Social Network Analysis* (SNA) [1]. Expectations within the networking community are that social insights could assist simplifying networking complexity and benefit the design of efficient protocols. Indeed, they have been shown to improve network functions such as routing decisions in opportunistic networks [2] [3], or content-caching strategies in wired networks [4].

Common denominator to all these efforts is the use of SNA-driven metrics for assessing the centrality, *i.e.*, significance, of individual network nodes, whether humans or servers. The computation of these metrics, however, typically demands *global* information about all network nodes and their interconnections. The distribution and maintenance of this information is problematic in large-scale networks. In certain environments it may not even be an option at all, as for example in the emerging self-organizing networks that lack centralized network management. A more realistic alternative for assessing node centrality may be based on its ego network (SNA term) or centered graph (graph-theoretic term) [5], *i.e.*, the subgraph involving itself, its 1-hop neighbors, and their interconnections. Nodes can acquire a *local approximation* of their centrality through *egocentric* measurements within their immediate locality; the degree centrality (DC) serves as the simplest estimate of that kind.

Computing egocentric[1] metrics is apparently of low complexity and, in fact, part of protocol implementations [2], [4]. Nevertheless, the capacity of these local approximations to substitute the globally computed sociocentric metrics is almost always taken for granted rather than evaluated. In physical topologies, it is the DC that is typically considered as the proper alias for the global centrality values [6], [7]. Over Internet router-level topologies, in particular, a systematic study of the effectiveness of local approximations is yet to come. On the one hand, it is this thread that our paper addresses; on the other, we seek to assess whether the typical finding of works on locally-determined approximations *i.e.*, the high positive correlation with the global metric, practically favors the effectiveness of network functions that rely on it.

We focus on node centrality metrics, which are the most commonly used in networking protocols. Besides the well known Betweenness Centrality (BC) metric, we study Conditional BC (CBC), an already introduced BC variant [8] particularly suited to networks with many-to-one data flow typologies. As in several cases it is the *order* of the metric values that matters rather than their *absolute values*, we first compare the sociocentric metrics, computed under global topological information, against their egocentric counterparts, computed locally over the nodes' ego networks, in terms of

This work has been supported by EINS, the Network of Excellence in Internet Science through EC's Grant Agreement FP7-ICT-288021.

[1]The terms egocentric and localized as well as sociocentric and global will be considered as synonyms and used interchangeably.

correlation-operations; namely the overlap of the k nodes exhibiting the top values of the considered counterparts and foremost, their corresponding *rank*-correlation. Then, contrary to previous works, we proceed to study *whether* a number of basic network functions as well as protocols, rely their effectiveness on rank-preserving localized centrality metrics.

A. Related work

Computing centrality values typically requires global topological information. BC computations in particular, can be accordingly carried out by exact [9] or approximation algorithms [10], [11]. The former explores the network using well-known algorithms (*i.e.*, BFS, Dijkstra) and then effectively aggregates the discovered paths circumventing the high complexity of the involved all-pairs shortest path problem; the later seek to provide accurate sampling methods that approximate the BC of each node based on number of the (single source) shortest paths originating from certain reference nodes or even restrict the considered node pairs to those that lie at most k hops away from each other [12].

On the contrary methods that rely on locally available information are increasingly sought for as they are better suited, if not the only alternative, to facilitate networking protocol operations; in light of the autonomic networking environments, the availability of networkwide[2] topological information is deemed non-realistic. Moreover, recent distributed-fashion approaches that leverage random walk strategies to reveal high BC [13], [14] or central nodes in a more general sense [15], do not qualify for the task as they require information gathering from the whole or part of the network topology, respectively.

Consequently, research efforts (in various disciplines) seek to devise localized counterparts of the global centrality metrics and study *whether* and *how well* do the former correlate with the latter. From a networking standpoint, localized counterparts have been proposed for the bridging [16] and closeness centrality [17]. In the first case the global BC factor of the original bridging centrality definition is replaced with the egocentric BC metric. In the second, the authors argue that the number of edges present in a small network neighborhood around a given node, accurately reflects its closeness to the rest of the network nodes provided that the considered network exhibits relatively small radius. Both works promote the high positive rank-correlation scores between the local and global counterpart over synthetic and real world networks (including router-level topologies for the second work), as evidence for the effective approximation of the latter but do not explore whether this is sufficient to drive networking operations.

Regarding BC, its localized approximation over real-world network topologies is typically provided by the node degree. This is based on the grounds of the observed linear behavior between the mean BC metric and the node degree [6]. The authors analyze the probability distribution of various quantities including betweenness centrality for different time snapshots of Internet maps. They observe that these distributions are characterized by scaling exponents stationary in time. Moreover, they argue that the nontrivial betweenness and degree correlation is due to the Internet hierarchical structure. However, note that these studies involve AS-level Internet maps that can exhibit different connectivity properties compared to the router-level topologies. Similar power-law behavior has been reported by Goh *et al.* [18] for the *load* distribution (which is only a betweenness centrality approximation) in synthetic scale-free networks The network's degree distribution follows a power law $p(k) \sim k^{\gamma}$ where $\gamma \in [2, \infty)$ while the load l has been shown to be distributed according to $P(l) \sim l^{\delta}$ with exponent δ. Based on numerical results, the authors have conjectured that the value of $\delta \simeq 2.2$ is independent of γ for the interval (2, 3] suggesting a "universal" behavior for the $load$ quantity in scale-free networks.

Finally, a number of social studies provide experimental evidence for positive correlation between sociocentric and egocentric BC over *small social* and *synthetic* networks [19], [20]. In [19], the evaluation includes social networks of sizes up to 217 nodes and in all cases high *Pearson's correlation* values are reported. The deviation of different nodes from the respective regression line is explained on the basis of their hub/bridging functionality within the network. In [20] similar positive conclusions are drawn for synthetic random graphs $\mathcal{RG}(N, p)$ of size $N \in [25, 500]$ and edge occurrence probability $p \in [0.1, 0.6]$. The authors observe stronger correlation when the network nodes have either very similar or very differentiated sociocentric BC scores.

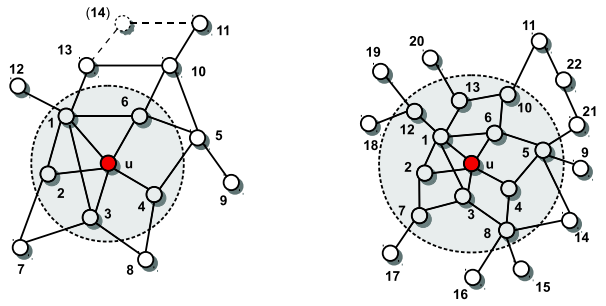
B. Our contribution

To exploit available information in a wider locality, we compute the egocentric betweenness variants over a generalized 2-hops ego network and analyze the corresponding time complexity and message overhead (Section II). Then we report on the BC *vs.* egoBC, BC *vs.* DC, CBC *vs.* egoCBC matching in terms of *correlation*-operations; we draw insights from synthetic topologies and then focus on (22 different ISPs') router-level topologies of sizes up to 80K nodes (Section III). Our synthetic networks yield significant rank-correlation that weakens as their size increases. In router-level topologies we measure high rank correlation (0.8-0.9) in almost all cases and even higher when the 2-hop neighbors are considered. On the contrary, the top-k overlap is found low prompting us to study whether the rank-preserving local metrics can *in practice* substitute the global ones. Our experiments with basic network functions such as the centrality-driven content search, show (Section IV) that high rank-correlation values between the centrality counterparts can hardly suggest the effectiveness of network functions utilizing the rank-preserving local counterparts.

II. Sociocentric *vs.* egocentric centrality metrics

The sociocentric metrics of betweenness centrality (BC) and the conditional betweenness centrality (CBC), along with

[2]The hierarchical structure of the Internet topologies prevents even a link state protocol from providing information of that scope.



a. Ego network of node u ($r = 1$). b. Ego network of node u ($r = 2$)

Fig. 1. a) Nodes 2, 3 and 4 contribute to $egoCBC(u; 11, 1) = 2$ with contributions 1/2, 1/2 and 1, respectively. b) Node 8 reaches the exit node 10 for destination node 11 through five different paths, two of which pass through node u, thus contributing 2/5 to $egoCBC(u; 11, 2)$.

their egocentric counterparts[3] are first presented. Then, the complexity savings achieved by the latter, are discussed.

A. Globally computed centrality indices

Consider an arbitrary node pair (s, t) over a connected undirected graph $G = (V, E)$. If σ_{st} is the number of shortest paths between s and t and $\sigma_{st}(u)$ those of them passing through node $u \neq t$, then the betweenness centrality of node u, equals $BC(u) = \sum_{\substack{s,t \in V \\ s<t}} \frac{\sigma_{st}(u)}{\sigma_{st}}$. Effectively, $BC(u)$ assesses the importance of a network node for serving information that flows over shortest paths in the network [21]. Whereas $BC(u)$ is an average over all network node pairs, the conditional betweenness centrality index (CBC), captures the topological centrality of node $u \neq t$ with respect to a *specific* destination node t [22] and is given by: $CBC(u; t) = \sum_{\substack{s \in V \\ s \neq t}} \frac{\sigma_{st}(u)}{\sigma_{st}}$, with $\sigma_{st}(s) = 0$. Therefore, CBC is particularly suited to settings where information is directed towards a particular node with discrete network functionality.

B. Locally computed centrality metrics - ego networks

Computing the above metrics requires information about the whole network topology and implies computational and message load overheads. In distributed settings, where nodes may be energy constrained or no explicit centralized network management be available, these computations are not favorable or not an option at all. Instead, techniques for *locally* assessing the centrality of network nodes can be borrowed from the SNA concepts. In the so-called *ego-network* structure of social studies the person we are interested in is referred to as the "ego" and its ego-network comprises itself together with those having an affiliation or friendship with it, known as "alters". Alters may as well share relations with each other (Fig. 1.a).

Hereafter, we generalize the ego-network definition to include nodes (alters) lying r hops away from u and the edges (links) between them. Formally, we can define the r^{th}-order ego network as follows. Let N_r^u be the set of nodes that form the r-hop neigborhood around u, *i.e.*, $N_r^u = \{n \in G : 1 \leq h(n, u) \leq r\}$, where $h(a, b)$ denotes the minimum hopcount between nodes a and b. The r^{th}-order ego network of node u is the graph $G_r^u = (V_r^u, E_r^u)$, where the set of nodes and edges are $V_r^u = N_r^u \cup \{u\}$ and $E_r^u = \{(i, j) \in E : i, j \in V_r^u\}$, respectively. For $r = 1$ the network G_1^u corresponds to the original ego network definition and consists of $|V_1^u| = DC(u) + 1$ nodes and $|E_1^u| = DC(u) + \mathcal{CC}(u) \cdot \binom{DC(u)}{2}$ edges, where $\mathcal{CC}(u)$ is the clustering coefficient [23] of node u and $DC(u)$ its degree centrality. Practically, values of $r > 2$ would tend to cancel the advantages that local ego-centered measurements induce.

Accordingly, BC metrics of a certain node can be defined with respect to its ego network. For the egocentrically measured betweenness centrality ($egoBC$) of node u, it suffices to apply the typical BC formula over the graph G_r^u : $egoBC(u; r) = BC(u)_{|V=V_r^u}$.

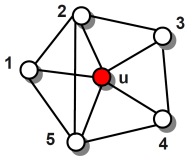


Fig. 2. Toy-example computation: $egoBC(u; 1)$= $10(1 - \frac{6}{10}) - [2(1 - \frac{1}{2}) + 2(1 - \frac{1}{3})] = \frac{5}{3}$

We further detail the computation formula of $egoBC(u, 1)$. In Fig. 2 the ego-node u is connected to $DC(u)$=5 first-neighbor nodes which are partially interconnected with each other. The maximum egoBC value that u can attain, when no links between its neighbors are present, equals $\binom{5}{2}$. From this value we need to subtract the number of pairs (like nodes 2-5 or 3-4) that share a direct link and therefore communicate without traversing u; their number equals the numerator of the clustering coefficient $\mathcal{CC}(u)$. Finally, we need to carefully account for those node pairs that share no direct link but are connected with multiple 2-hop paths, such as nodes 2 and 4 that are connected via the paths 2-3-4, 2-u-4 and 2-5-4. The idea is that only one of those paths will cross u; thus, we need to discount the original contribution (*i.e.*, unit) of each non-directly connected node pair $i, j \in N_1^u$ by as much as the inverse of the number of competing 2-hop paths connecting i and j. This number corresponds to the element (i, j) of matrix A^2, where A is the adjacency matrix for G_1^u. Summing up, we derive the expression for the first-order egocentric betweenness centrality in Eq. 1.

The egocentric counterpart of conditional betweenness centrality ($egoCBC$), on the other hand, is less straightforward. For each ego network and for a given destination node t, we need to identify the set of *exit* nodes $e_r(u; t) = \{t' \in N_r^u : h(u, t') + h(t', t) = h(u, t)\}$, *i.e.*, all nodes r hops away from the ego node u that lie on the shortest path(s) from u to t. This set is effectively the *projection* of the remote node t on the local ego network and may be a singleton but never the null set. In Fig. 1, for example, we have $e_1(u; 11) = \{6\}$, $e_1(u; 9) = \{4, 6\}$ for the G_1^u (left graph) and $e_2(u; 11) = \{10\}$, $e_2(u; 14) = \{5, 8\}$ for the G_2^u (right graph).

[3]The degree centrality (DC) is by default an egocentric centrality metric.

$$egoBC(u;1) = \begin{cases} \binom{DC(u)}{2}(1-\mathcal{CC}(u)) - \sum_{i,j \in N_1^u : h(i,j)=2}(1-\frac{1}{A^2_{i,j}}) & \text{if } DC(u) > 1 \\ 0 & \text{if } DC(u) = 1 \end{cases} \quad (1)$$

For each node $s \in G_r^u$, we need to calculate the fraction of shortest paths from s towards *any* of the nodes in $e_r(u;t)$ that traverse the ego node. Thus the egocentric variant of CBC is given by

$$egoCBC(u;t,r) = \sum_{\substack{s \in V_r^u \\ t' \in e_r(u;t)}} \frac{\sigma_{st'}(u)}{\sigma_{st'}} \mathbf{1}_{\{h(s,t') \leq h(s,l),\ l \in e_r(u;t)\}} \quad (2)$$

Again, in Fig. 1a, node 4 contributes to the $egoCBC(u,11)$ value since its shortest path to the single exit node 6 traverses u, although it has a shorter path to node 11, via nodes {5,10} that lie outside the ego network. Likewise, its contribution is a full unit, rather than $1/2$, since the second shortest path to node 6 passes through node 5, a node outside the ego network of u. This is the price egocentric metrics pay for being agnostic of the world outside their r-neighborhood.

Although, the definitions of both ego- and sociocentric metrics are valid under weighted and unweighted graphs, we focus on the latter ones. The way link weights affect the correlation operations is clearly worth of a separate study.

C. Complexity comparison of betweenness counterparts

We briefly discuss how the two types of metrics compare in terms of message overhead and time complexity required for their computation. Message overhead is measured in messages times number of edges they have to travel. In both cases, we can distinguish two metric computation phases: the collection of topological information and the execution of computations.

Sociocentric computation of centrality. The network nodes need to collect global information about the overall network topology; hence, each one of the $|V|$ network nodes has to inform the other $|V|-1$ about its neighbors. This generally requires $O(|E_f|)$ message copies and $O(D)$ time steps for each node's message, where D is the network diameter and $|E_f|$ the number of edges in the flooding subgraph. In the best case, the flooding takes place over the nodes' spanning trees, hence the message overhead is $O(|V|-1)$. For the distribution of one round of messages by all nodes, the overhead becomes $O(|V|^2)$; the time remains $O(D)$ assuming that the process evolves in parallel. With knowledge of the global topology, each node can compute the BC values of all other nodes in the network. An efficient way to do this is to invoke Brandes' algorithm [9], featuring $O(|V|\cdot|E|)$ complexity for unweighted graphs and $O\big((|E|+|V|)\cdot|V|log|V|\big)$ complexity for weighted graphs. Interestingly, the CBC values of each network node with respect to all other network nodes, emerge as intermediate results of Brandes' algorithm for the BC computation [4].

Egocentric computation of centrality. Intuitively, the egocentric variants save complexity. The message overhead over the whole network is $O(2 \cdot |E|)$ for the ego network with $r = 1$ and $O(2 \cdot d_{max}|E|)$ for the ego network with $r = 2$, where d_{max} is the maximum node degree; for dense graphs this overhead becomes $O(|V|^2)$. The time required for the distribution of information is of no concern, $O(1)$. The egoBC and egoCBC computation for $r = 1$ can be carried out as in [20]. The computation involves a multiplication of an $O(d_{max})$-size square matrix and trivial condition checks. For $r = 2$, we can employ [9] replacing $|V|$ with d^2_{max}. Finally, DC is considered to be immediately available to every node.

TABLE I
COMPLEXITY COMPARISON OF SOCIO- *vs.* EGO-CENTRIC METRICS

Metric	Time complexity	Message overhead
BC	$O(\|V\|^3)$	$O(D \cdot \|V\|)$
egoBC (**r=1**)	$O(d^3_{max})$	$O(2 \cdot \|E\|)$
egoBC (**r=2**)	$O(d^4_{max})$	$O(2 \cdot d_{max} \cdot \|E\|)$
CBC	$O(\|V\|^3)$	$O(D \cdot \|V\|)$
egoCBC(**r=1**)	$O(d^3_{max})$	$O(2 \cdot \|E\|)$
egoCBC(**r=2**)	$O(d^4_{max})$	$O(2 \cdot d_{max} \cdot \|E\|)$
DC	$O(1)$	–

As expected, since d_{max} is typically much smaller than $|V|$, the use of local metrics bears apparent computational benefits. The question of whether these metrics correlate well with the sociocentric ones is considered next.

III. EXPERIMENTAL CORRELATION OPERATIONS STUDY BETWEEN SOCIO- AND EGOCENTRIC CENTRALITY METRICS

A. Correlation coefficients

In comparing the ego- with sociocentric metrics, we are mostly concerned with their *rank* correlation. The underlying remark is that in protocol implementations that seek to utilize highly central nodes, we usually care more about the way the metric *ranks* the network nodes rather than their absolute metric values. For example, this is how a degree-based search scheme explores unstructured p2p networks [7] or the egoBC metric is used in DTN routing protocols [2], [3]. Alternatively, we may need to utilize only a small number of k nodes that exhibit the top centrality values [22] and therefore, we include in our study the ovelap scores between the top-k nodes determined with ego- and sociocentric counterparts, respectively. We capture the rank correlation in the non-parametric Spearman measure of correlation, ρ, which assesses

[4]The algorithm in [9] effectively visits successively each node $u \in V$ and runs augmented versions of shortest path algorithms. By the end of each run, the algorithm has computed the $|V|-1$ $CBC(v;u)$ values, $v \in V$; while the $|V|$ $BC(v)$ values result from iteratively summing these values as the algorithm visits all network nodes $u \in V$.

how monotonic is the relationship between the ranks of the two centrality variables and is computed as follows:

$$\rho = 1 - \frac{6 \sum\limits_{u \in V} (r_s(u) - r_e(u))^2}{|V|(|V|^2 - 1)} \quad (3)$$

where $r_s(u)$ and $r_e(u)$ are the ranks of each graph node when ordered according to the sociocentric and egocentric definition of the metrics, respectively. For the sake of a comprehensive study we complete the presented results with the well-known linear Pearson correlation. The r_{Prs} coefficient as well assesses a straight-line relationship between the two variables but now the calculation is based on the actual data values. Both ρ and r_{Prs} lie in [-1,1]. If being close to 1 we can only infer significant correlation between the considered variables. For the pairs of the socio- and ego- betweenness variants $(sB(u), eB(u))$ of each node $u \in V$, the Pearson r is given by the equation:

$$r = \frac{\sum\limits_{u \in V} (sB(u) - \overline{sB})(eB(u) - \overline{eB})}{\sqrt{\sum\limits_{u \in V} (sB(u) - \overline{sB})^2} \sqrt{\sum\limits_{u \in V} (eB(u) - \overline{eB})^2}} \quad (4)$$

When the involved parameters vary along the node location (*i.e.*, CBC values), we present the correlation averages along with the 95% confidence intervals, estimated over an at least 6% of the total locations, sample.

B. Network topologies

We have placed emphasis on the experimentation study with the real-world intradomain Internet topologies. Nevertheless, synthetic graph models with distinct structural properties such as the two-dimensional rectangular grid have been as well employed to complete our study. The router-level ISP topolgies, on the other hand, do not have the predictable structure and properties of the synthetic topologies and may differ substantially one from another, typically sizing up to a few thousands of nodes. Yet it is over such topolgies that networking protocols will seek to rely their operation on localized centrality metrics [22]. Three sets of router-level topologies discovered by different techniques have been employed:

- `mrinfo` *topologies:* The dataset we consider [24] includes topology data from 850 distinct snapshots of 14 different AS topologies, corresponding to Tier-1, Transit and Stub ISPs [25]. The data were collected daily during the period 2004-08 with the help of a multicast discovering tool called `mrinfo`, which circumvents the complexity and inaccuracy of more conventional measurement tools. Herein we present and discuss results from a representative subset of the datasets (Tier-1 and Transit ISP networks) exhibiting adequate variance in size, diameter, and connectivity degree statistics.
- *Rocketfuel topologies:* The Rocketfuel technique [26] has been shown to collect high-fidelity router-level maps of ISPs and therefore has been widely used despite its old, around 2002, publication. The considered dataset [27] includes measurements from 800 vantage points serving as `traceroute` sources. A number of innovative techniques such as BGP directed probing, use of DNS names and IP identifiers have been applied to reduce the number of probes, identify what part of the trace belongs to a certain ISP and discover the different interface IP addresses that belong to the same router (*i.e.*, alias resolution), respectively.
- *CAIDA topologies:* The third and most recent dataset [28] we have used in our experiments was collected during Oct-Nov 2011 by CAIDA performing `traceroute` probes to randomly-chosen destinations from 54 monitors located in 29 countries worldwide. We have processed the router-level raw data files derived from tools (*i.e.*, MIDAR and iffinder [28]) that were run over approximately two million IPv4 addresses and achieved the highest aliases-resolution confidence; parsing the provided separate file that heuristically assigns an AS to each node found, we have determined the router-to-AS ownership and subsequently have extracted out of the large raw data files the topologies of the nodes operated by certain ASes. Our effort was to discover the largest ISP networks present in the dataset to further broaden our experimentation topology set.

With all three datasets we avail a rich experimentation basis of a diverse Internet topologies' set that can limit the effect of any measurement technique errors and provide our results with extra credibility.

TABLE II
CORRELATION STUDY BETWEEN BC AND EGOBC ON GRID NETWORKS

Grid size	Diameter / Mean degree	Spearman ρ	
		ego-network **(r=1)**	ego-network **(r=2)**
5x5	8 / 3.200	0.9195	0.9679
10x10	18 / 3.600	0.8400	0.9556
15x15	28 / 3.733	0.7510	0.9017
20x20	38 / 3.800	0.6802	0.8459
50x50	98 / 3.920	0.2429	0.2942
60x8	66 / 3.717	0.5735	0.6336
90x8	96 / 3.728	0.5390	0.5870
150x8	156 / 3.737	0.4584	0.4181
400x8	406 / 3.745	0.1633	0.2213

C. Experimental results

We choose to spend more of our effort experimenting on the BC-egoBC and BC-DC correlation debates that are expected to attract more interest (see relevant applications in the Section IV) as opposed to the limited scope of CBC. Note also that in router-level Internet topologies the BC-DC correlation has been so far taken for granted rather than experimentally shown.

*egoBC vs. BC :*In the grid topologies, ego networks have fixed size depending on their relative position, *i.e.*, corner, side, or internal nodes. The ego networks are star networks of 3, 4 and 5 nodes, respectively, for $r = 1$; they may have size up to 6, 9, and 12 nodes, respectively, for $r = 2$. Thus, the egoBC index may only exhibit three values (*i.e.*, 1, 3 and 6) with respect to the node's location when $r = 1$ In case of $r = 2$

the egoBC cannot exceed the value of 28 (*i.e.*, when the ego network size reaches its upper bound). Table II suggests that the rank correlation values decrease monotonically with the grid size.

As one or both grid dimensions grow larger, the number of shortest paths between any node pair grows exponentially, resulting in a richer spectrum of BC metric values over the grid node population. On the other hand, the possible egoBC values remain the same; only the distribution of grid nodes over these values changes (see Fig. 3). The same behavior and reasoning holds for another regular topology, the line network, where the corresponding rank correlation may be as high as 1 for a line of a few nodes while fast decreasing as its size grows larger.

Our findings for the real-world ISP topologies are listed in Table III. Even with measurements within the first-order ego network, there is high positive rank correlation between BC and egoBC. Yet some ISP networks (*e.g.*, AT&T and Sprint) show clearly stronger correlation than others (*e.g.*, China Telecom). On the other hand, the Pearson linear correlation coefficient suggests looser yet positive association between the two variants (in almost all datasets). If nodes are willing to tolerate the extra overhead related to computing egoBC in the second-order ego network (see II-C), the correlation values for both coefficients become even higher but also more similar with each other, despite the differentiation in network size. The general structural characteristics of the considered ISP topologies differ from the grid topologies. No regularities are reported; their diameter and clustering coefficient attain many different values as their size scales. Provably there is enough asymmetry throughout the topology to yield a wide range of BC and egoBC values and favor high correlation values between the two. There is one notable exception from this rule, the Level-3 ISP topology (dataset 12,13) of the `mrinfo` dataset explored later in Section III-D. Finally, note that the above results are also relevant to the localized bridging centrality metric [16] which employs the egoBC as an alias to the BC metric; it follows that the two bridging centrality counterparts are higly correlated in these topologies.

DC vs. BC: We examine the DC values which can be immediately available at each node, as a candidate to locally approximate the corresponding nodes' BC. In the considered synthetic topologies we have observed the same correlation (both Spearman and Pearson) degradation with the network size, explained along the aforementioned BC-spectrum arguments. In the router-level topologies (Table III) we find high Pearson and even higher Spearman correlation over a broad range of ISPs although consistently lower than the corresponding egoBC *vs.* BC one, at least for the values of the latter coefficient. Our results actually extend the previously reported [6] or inferred [18] (based on the scaling laws of the load *i.e.*, a BC approximation) DC-BC correlation over AS-level topologies and synthetic scale-free networks, respectively, to the intra-domain router-level topologies. Moreover, note that the observed high rank correlation between egoBC and BC can be further justified on the grounds of the high DC-BC Spearman values. Indeed, the router-level topologies exhibit very small mean clustering coefficient values and therefore, according to eq. 1 the egoBC metric attains almost identical values to the DC ones. This is depicted in the scatterplot of Fig. 4 where the egoBC-DC correlation appears to scale linearly with the mean clustering coefficients, especially in the CAIDA dataset.

egoCBC vs. CBC : We devote a different set of experiments to the assessment of the ego variant of the CBC metric (r=1). The high ρ values of Table IV suggest significant positive rank correlation in all considered ISP topologies. Especially for the outlier case of the `mrinfo` Level-3 networks the correlation turns out to be considerably increased, compared to the one between the BC variants. Intuitively, the correlation of CBC with egoCBC values is expected to be higher than the BC *vs.* ego-BC counterpart; by neglecting the world outside of the ego network, the egoBC inaccuracies (compared to the globally determined BC) may arise anywhere across the whole network. On the contrary, the $egoCBC(u;t,r)$ considers only the paths that lead to the target t, somehow focusing on an angle that encompasses t; thus, it may differ from the CBC(u;t) view only across that certain angle.

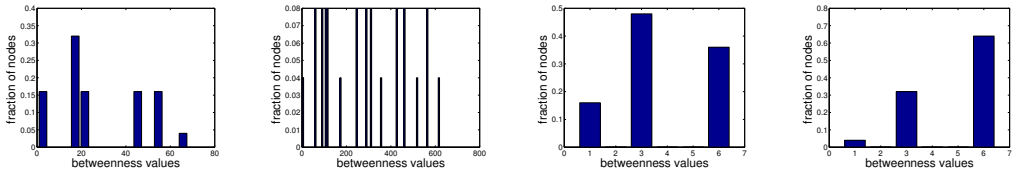
top-k nodes overlap : Our final experiment seeks to assess the extend to which the nodes with the top-k values of localized centrality metrics coincide with the corresponding ones determined with respect to the global BC. It is interesting to observe in Table V that the top-k nodes identified by the different localized counterparts exhibit relatively low overlap, at least for the first-order ego network measurements, with the globally top central ones. This does not actually contradicts our previous results since the correlation coefficients are determined over all network nodes rather than a subset of cardinality k. Clearly, the observed high rank-correlation is mainly due to nodes of lower rank; for instance, those with already zero values for both centrality counterparts (*i.e.*, DC=1) have been reported to drastically contribute to the high egoBC-BC correlation [20]. Along this thread, our comprehensive study of the following subsection shows that the actual association between the two metric variants is not determined solely by the degree distribution.

Finally, the low top-k overlap scores a) help us infer that the protocols utilizing a handful of top central nodes through localized approximations, are expected to perform poorly. b) serve as a warning sign regarding to what the high coefficients can reveal about the practical implications of rank-preserving localized centrality metrics (see Section IV).

D. Pathologies in the mrinfo Level-3 ISP snapshots

We now come back to the `mrinfo` dataset where the Level-3 ISP topologies exhibit an exceptional behavior with respect to the outcome of the studied BC-egoBC correlation (see Table III).

Neither the relatively extreme values of the network diameter, nor the considerably higher mean degree of Level-3 topologies can justify the poor ρ values or even more, the negative r_{Prs} ones. To further elaborate on this, we studied



a. BC (5x5) b. BC (10x10) c. egoBC (5x5) d. egoBC (10x10)

Fig. 3. Probability distribution of BC and egoBC values for scaling size of a grid network.

TABLE III
CORRELATION STUDY BETWEEN BC-EGOBC AND BC-DC ON INTRA-DOMAIN ISP TOPOLOGIES

DataSet	ID	ISP(AS number)	<$\mathcal{CC}$>	Diameter	Size	<degree>	**BC *vs.* ego-BC**				**BC *vs.* DC**	
							Spearman ρ		Pearson r_{Prs}			
							ego-net. **r=1**	ego-net. **r=2**	ego-net. **r=1**	ego-net. **r=2**	Spearman ρ	Pearson r_{Prs}
	36	Global Crossing(3549)	0.546	10	76	3.71	0.9648	0.9806	0.6720	0.9197	0.8836	0.6564
	35	-//-	0.479	9	100	3.78	0.9690	0.9853	0.7029	0.9255	0.8506	0.6714
	33	NTTC-Gin(2914)	0.307	11	180	3.53	0.9209	0.9565	0.7479	0.8561	0.8180	0.6664
	21	Sprint(1239)	0.298	12	216	3.07	0.9718	0.9812	0.7470	0.8557	0.8824	0.6780
	13	Level-3(3356)	0.169	25	378	4.49	0.2708	0.9393	-0.0918	0.7982	0.1953	-0.0813
m	12	-//-	0.149	28	436	4.98	0.2055	0.9381	-0.1217	0.7392	0.1696	-0.1128
r	20	Sprint(1239)	0.287	16	528	3.13	0.9866	0.9928	0.5805	0.8488	0.8543	0.6815
i	9	-//-	0.251	13	741	3.29	0.9901	0.9930	0.7149	0.8622	0.8568	0.7926
n	40	JanetUK(786)	0.132	14	336	2.69	0.9714	0.9825	0.8049	0.9180	0.8855	0.7706
f	45	Iunet(1267)	0.246	11	598	3.88	0.8506	0.9468	0.8887	0.9688	0.8309	0.7447
o	38	-//-	0.231	12	645	3.75	0.8790	0.9516	0.9094	0.9568	0.8549	0.7708
	39	-//-	0.038	13	711	3.45	0.9470	0.9826	0.5354	0.9536	0.9495	0.5405
	44	Telecom Italia(3269)	0.037	13	995	3.65	0.7950	0.9828	0.3362	0.8699	0.7733	0.4852
	50	TeleDanmark(3292)	0.058	15	1240	3.06	0.9569	0.9738	0.5475	0.9025	0.9388	0.5538
R	60	VSNL(4755)	0.263	6	41	3.32	0.9909	0.9971	0.7286	0.9603	0.8740	0.7842
O	61	Ebone(1755)	0.115	13	295	3.68	0.9736	0.9860	0.6856	0.8895	0.9443	0.7457
C	62	Tiscali(3257)	0.028	14	411	3.18	0.9522	0.9659	0.6073	0.9281	0.9464	0.7103
K	63	Exodus(3967)	0.273	14	353	4.65	0.9125	0.9792	0.6100	0.9061	0.8204	0.6241
E	64	Telstra (1221)	0.015	15	2515	2.42	0.9990	0.9990	0.3336	0.7565	0.9783	0.5172
T	65	Sprint(1239)	0.022	13	7303	2.71	0.9980	0.9990	0.4770	0.7977	0.9562	0.6537
	66	Level-3(3356)	0.097	10	1620	8.32	0.9841	0.9923	0.6346	0.9075	0.9655	0.7045
F	67	AT&T(7018)	0.005	14	9418	2.48	0.9988	0.9994	0.3388	0.5302	0.9882	0.4483
L	68	Verio (2914)	0.071	15	4607	3.28	0.9904	0.9969	0.4729	0.8044	0.9315	0.6718
	70	UUNet (701)	0.012	15	18281	2.77	0.9841	0.9886	0.5430	0.8752	0.9694	0.7544
C	71	COGENT/PSI(174)	0.062	32	14413	3.09	0.9638	0.9599	0.7272	0.9354	0.8940	0.8791
A	72	LDComNet(15557)	0.021	40	6598	2.47	0.9674	0.9245	0.3782	0.7676	0.9479	0.6634
I	73	TeliaNet(1299)	0.037	13	3820	3.08	0.9593	0.9764	0.9176	0.9628	0.9047	0.9594
D	74	ChinaTelecom(4134)	0.083	19	81121	3.97	0.8324	0.8986	0.7861	0.9714	0.7370	0.8795
A	75	FUSE-NET(6181)	0.018	10	1831	2.38	0.9903	0.9763	0.6205	0.8574	0.9536	0.7445
	76	JanetUK(786)	0.031	24	2259	2.26	0.9819	0.9834	0.4444	0.8506	0.9450	0.5765

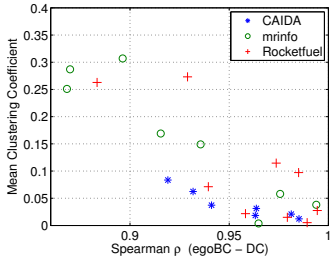


Fig. 4. Linear relations between the egoBC-DC correlation and <CC> values.

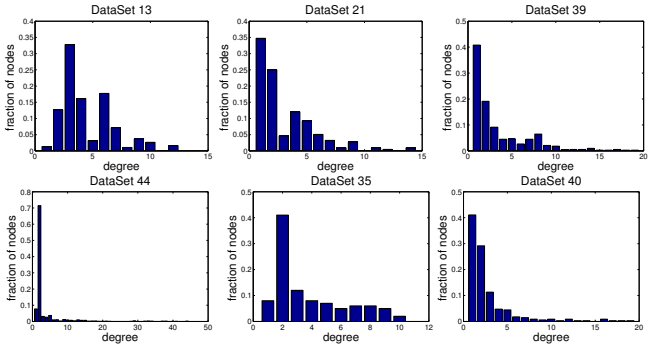


Fig. 5. Indicative set of the mrinfo ISP topologies' degree distribution

the degree distribution of these topologies that were found to exhibit an unusual to the whole dataset, bimodal-like shape with two distinct high peaks at values 3 and 6 and little mass at value 1 (see Fig. 5). This finding seems consistent with the remark about the role of nodes with degree one in [20]. Nodes with degree 1 have zero values of both BC and egoBC. Therefore, a network availing a large percentage of such nodes has higher chances for an overall higher correlation coefficient than those with fewer nodes. To test further this hypothesis, we generated multiple random graphs [29] feeding the generator with the degree distribution of the Level-3 outliers. Precisely, for each one of the Level-3 datasets 16, 18, 27 and 28 we have generated a couple (*i.e.*, named a and b, respectively) of new topologies characterized by the same degree distribution as the Level-3 original ones. In

TABLE IV
CORRELATION BETWEEN CBC AND EGO-CBC (R=1) ON ISP TOPOLOGIES

Dataset ID	Spearman ρ	95% Confidence interval
36	0.9568	0.008
35	0.9489	0.013
33	0.9554	0.003
21	0.9824	0.002
13	0.7336	0.007
12	0.7035	0.005
20	0.9847	0.003
9	0.9884	0.002
40	0.9819	0.001
45	0.7825	0.033
38	0.8062	0.022
39	0.9370	0.016
44	0.9902	0.001
50	0.9739	0.009
60	0.9430	0.076
61	0.8423	0.027
62	0.9321	0.016
63	0.7641	0.023
75	0.9961	0.005
76	0.9853	0.002

Table VI we report the properties of each topology along with the measured BC-egoBC rank correlation. The resulting high correlation between BC and egoBC-driven node rankings of the generated topologies (*i.e.*, 0.84 to 0.95) implies that the actual association between the two metric variants is not determined solely by the degree distribution. Note that the rank correlation for Level-3 topology snapshots (back in Table III) improves significantly when egoBC is computed within the second-order ego network.

Following the remark about nodes with degree one we present in Table VI the α_{EQ} percentage of nodes (for both Level-3 ISP and the corresponding generated network) that exhibit equal BC and egoBC values. The relative high α_{EQ} (7.2-9.1%) of the Level-3 topologies is not sufficient to result in high positive correlation between BC and egoBC values. Taking one step further, we measure the percentage α_{LE} of network nodes that attain lower BC values than egoBC. Interestingly, the original Level-3 exhibit values of α_{LE} as high as 23% whereas the corresponding generated topologies avail no such nodes. Contrary to intuition it seems that the considered Level-3 topologies avail a critical mass of nodes for which the transition from the restricted egocentric view to the global topology approach results in the decrease of the corresponding centrality values. Clearly, such nodes have negative contribution to the measured BC-egoBC rank correlation.

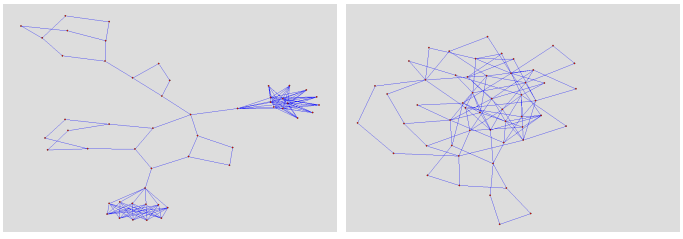


Fig. 6. Vizualization of an original mrinfo Level-3 snapshot (left) and a generated topology with the same degree distribution (right).

Fig. 7. Configurations of topology's cluster endings that result in egoBC values of certain nodes exceeding their BC ones.

To shed some light on the topology configuration that causes this counterintuitive result, we visualize in Figure 6 an original Level-3 topology together with its corresponding generated one. The Level-3 snapshot seems to consist of a lengthy backbone on which two cluster endings are attached. It is the configuration of the links among the cluster nodes that makes some of them unreachable from shortest paths that stem from network nodes outside the current cluster. The corresponding high α_{LE} percentage is due to a significant portion of such cluster nodes; node like 5 in Fig. 7(left) or 5, 6 and 7 in Fig. 7(right) are the ones that fall in this category and therefore, exhibit higher ego- than socio- BC values. In particular node 5 in Fig. 7a has an egoBC value equal to $\binom{d(5)}{2} = \binom{4}{2} = 6$ while its socio- variant is determined by all possible pairs between nodes 1, 2, 3 and 4; one out of the two shortest paths linking those pairs traverse node 5 *i.e.*, $BC(5) = 1/2 \cdot \binom{4}{2} = 3$. Consequently, the higher the number of Level-3 clusters, the higher the number of nodes that impede the BC variants' high rank correlation. On the contrary the Dataset 18a generated topology (with the same degree distribution as 18) forms no cluster and furthermore, exhibits zero values for the α_{LE} percentage. This characteristic turns out to provide the topology with high correlation between the centrality variants even if the percentage α_{EQ} remains significantly low. Finally, note that the Level-3 snapshot of the Rocketfuel dataset (*i.e.*, id 66) published about 8 years earlier than the `mrinfo` one, does not exhibit the above problematic clustered configurations.

IV. PRACTICAL UTILITY OF LOCAL CENTRALITY METRICS

We outline examples, where locally measured centrality metrics are employed, encouraged by the high corresponding correlation, to substitute the global ones and assist with critical network operations. Our aim is to assess whether such options can actually be effective in networking practice.

A. *Identifying vulnerable network locations: a resilience study*

First, we elaborate on a rather theoretical scenario *i.e.*, sequentially selecting crucial network nodes and studying the effect of having them "shut-down"(removed) on the router-level connectivity structure. The node selection is carried out on the basis of the each time top value of a global(*i.e.*, BC) or alternatively local(*i.e.*, egoBC, DC) centrality metric. Typically, we are interested in measuring the resulted increase of the average path length and the clustering coefficient variations. However, we observe that the ISP topologies bear

TABLE V
OVERLAP(%) BETWEEN NODES WITH THE TOP-k LOCALIZED CENTRALITY AND BC VALUES

	k=10			k=30		
Dataset ID	egoBC(r=1)	egoBC(r=2)	DC	egoBC(r=1)	egoBC(r=2)	DC
50	30.0	70.0	30.0	10.0	60.0	10.0
63	10.0	60.0	10.0	0.0	30.0	0.0
67	0.0	10.0	0.0	0.0	30.0	0.0
70	0.0	90.0	0.0	36.7	76.7	43.3
71	40.0	90.0	40.0	56.7	80.0	60.0
72	40.0	50.0	40.0	50.0	60.0	50.0

TABLE VI
PROPERTIES OF ORIGINAL MRINFO LEVEL-3 (*bold*) AND GENERATED TOPOLOGIES WITH THE SAME DEGREE DISTRIBUTION

DataSet	ISP (AS number)	$<\mathcal{CC}>$	Diameter	Size	<degree>	variance	Spearman ρ	α_{EQ} (%)	α_{LE} (%)
16	**Level-3 (3356)**	**0.1307**	**28**	**431**	**5.00**	**6.57**	**0.1563**	**7.19**	**22.04**
16a	*generated*	0.0244	8	-//-	4.95	6.34	0.9580	0.93	0
16b	*generated*	0.0094	8	-//-	4.93	6.08	0.9440	0.92	0
18	**Level-3 (3356)**	**0**	**12**	**52**	**4.54**	**4.61**	**0.0183**	**0**	**23.08**
18a	*generated*	0.0516	6	-//-	4.23	3.79	0.8495	1.92	0
18b	*generated*	0.1034	6	-//-	4.15	3.70	0.9462	5.76	0
27	**Level-3 (3356)**	**0.1809**	**24**	**339**	**3.98**	**3.47**	**0.4130**	**9.14**	**15.92**
27a	*generated*	0.0102	10	-//-	3.94	3.22	0.9053	1.48	0
27b	*generated*	0.0118	9	-//-	3.96	3.29	0.9281	1.18	0
28	**Level-3 (3356)**	**0.1749**	**24**	**349**	**4.10**	**3.86**	**0.3522**	**8.88**	**16.05**
28a	*generated*	0.0074	10	-//-	4.09	3.77	0.9069	1.14	0
28b	*generated*	0.0169	10	-//-	4.03	3.49	0.9229	2.00	0

TABLE VII
NUMBER OF NODES TO BE REMOVED FOR NETWORK FRAGMENTATION

Node removal criterion	Dataset's fragmentation number ID 50	ID 64	ID 71	ID 72
top BC value	2	3	1	1
top egoBC(r=1) value	2	1	1	2
top egoBC(r=2) value	2	1	1	1
top DC value	2	1	1	2

very few node removals (encountered by the fragmentation number in Table VII) until they become disconnected and thus, hindering the detailed assessment of the localized metrics' capacity to identify the same critical nodes as the global do.

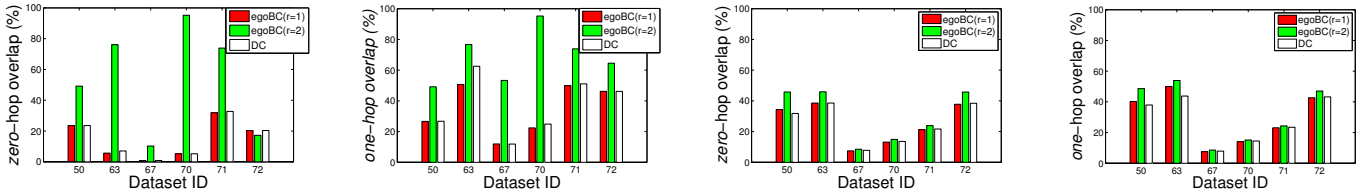
Two remarks coming out of this study are a) the egoBC and DC appear to be equivalent in terms of locating nodes that hold the network together as opposed to the BC metric which may render different ones as important; as earlier discussed, it is the vanishing clustering coefficient values that result in the egoBC values closely approximating the DC ones. and b) the fact that the attack on high DC/egoBC nodes results in the fragmentation of the ISP topologies is in good agreement with previous studies that model the router-level topologies as scale-free graphs and argue about their robustness to randomly chosen nodes' removal, likely to take out those with small degree. On the contrary, the removal of a few key-hubs splinters the topology into groups of isolated routers [30].

B. A local-centrality-driven navigation scheme

We now explore whether a practical navigation scheme can benefit from local centrality metrics. To replace the wasteful broadcasting of queries in search of a target data file in unstructured power-law P2P networks, it has been shown [7] that a high-degree seeking strategy is more efficient than a random-walk strategy. The former, at each step, passes a single query message to a neighboring node of higher degree exploiting the availability of many network links pointing to high degree nodes. Following the thread, the work in [31] proposes a local-scope BC (LBC) to drive the search and shows that the difference between the high-degree and high-LBC search increases with the network edge weights heterogeneity.

We have implemented a similar navigation scheme that crawls the network following a MAX or MAX-MIN pattern with respect to node centrality; each time, the crawler moves to the neighbor with the maximum(minimum) centrality out of those that exhibit higher(lower) values than the current node, utilizing a self-avoiding path. We randomly select 20% of the network nodes as starting points and execute 10 runs (*i.e.*, crawlings) for each one but with no destinations predetermined. Effectively, we seek to compare the navigation patterns and final locations achieved by using the different centrality variants as drivers. α-hops overlap measures the percentage of the final locations lying within α hops away from those the global metric yields. Zero-hop overlap clearly refers to the destinations' exact matching. Regarding the final locations, figs. 8a-d) show that the local metrics (r=1) can hardly be effective aliases for the global ones. In Table VIII the crawler, driven by local metrics, is measured to consistently take, up to 2.3 on average, less hops than when it is BC-driven. It also reaches multiple unique final locations (*i.e.*, UFL) compared to the few ones where it ends-up driven by BC; similar results are obtained for the MAX-MIN pattern. These findings clearly suggest that the crawler driven by local centrality metrics fails to identify the same navigation pattern *i.e.*, sequence of central nodes the global BC identifies.

In view of the high correlation between the involved met-



a. MAX navigation pattern b. MAX navigation pattern c. MAX-MIN navigation pattern d. MAX-MIN navigation pattern

Fig. 8. Performance of local-information-based search scheme in terms of $zero$- and one-hop overlap with the BC-determined final locations.

ID	50		63		67		70		72	
	<hopcount>	UFL	<hopcount>	UFL	<hopcount>	UFL	<hopcount>	UFL	<hopcount>	UFL
BC	4.6573 ± 1.6761	9	3.5493 ± 1.2738	4	4.3864 ± 0.9548	14	4.2232 ± 1.0126	21	4.9333 ± 2.4374	51
egoBC(r=1)	2.3476 ± 0.8970	78	2.7042 ± 1.1513	18	2.6976 ± 0.9642	179	2.8288 ± 1.1728	249	2.7961 ± 1.2627	330
egoBC(r=2)	4.0677 ± 1.6508	18	2.7042 ± 1.1513	4	2.9644 ± 0.4203	47	4.3679 ± 0.9112	7	4.1402 ± 1.7120	89
DC	2.3310 ± 0.8845	79	2.6930 ± 1.1639	18	2.7162 ± 0.9711	174	2.7257 ± 1.0175	235	2.7936 ± 1.2565	332

TABLE VIII
MEAN HOPCOUNT AND UNIQUE FINAL LOCATIONS (UFL) FOR THE MAX PATTERN CENTRALITY-DRIVEN NAVIGATION SCHEME

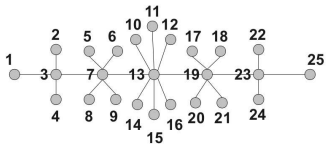


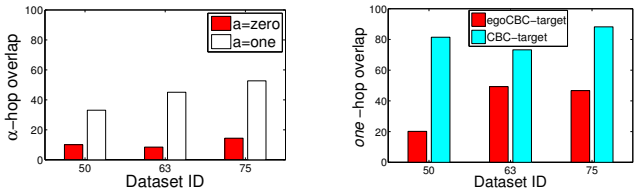
Fig. 9. Toy topology with egoBC-BC and DC-BC *perfect* rank correlation.

rics, we have tried to shed some light on the somewhat counterintuitive, poor navigation performance. We have built a toy topology (Fig. 9 where the rank correlation between the considered counterparts is perfect (*i.e.*, $\rho = 1$). Removing sequentially the nodes 5,6,17 and 18, we measure the rank-correlation reducing from 1 to 0.9953 while the $zero$-hop overlap for the MAX pattern drastically diminishes from 100% to 61.90%. Clearly, the numerical summaries provided by the correlation coefficients fail to capture/reflect in micro-level the relative significance of each node which heavily determines the scheme's performance.

C. A local-centrality-driven search scheme

As the conditional centrality metrics involve a target node, we are enabled to compare CBC and egoCBC essentially over a (content) search scheme. For each starting location we randomly select a target node and seek to reach it utilizing a MAX search pattern. Fig. 10.a shows low ovelap between the final locations achieved by the two counterparts while the hopcount to the final location is again measured consistently lower (*i.e.*, 0.3 to 1.5 hops) for the egoCBC case. Driven by the local variant the search fails to track closely the route that the global one induces.

In terms of the one-hop overlap between the achieved final locations and the targets (*i.e.*, hit-rates), which is also reported in [7], figure 10.b shows that the egoCBC-driven search indeed hits significantly less targets than the CBC-driven does. The number of targets reached by the local CBC variant is in good agreement with the discovered P2P nodes in [7], using DC as a driver; there, a networkwide path of the query may cumulatively hit up to 50% of either the targets or their first neighbors. On the contrary, we obtain high values when the global metric drives the MAX search pattern, since the closer a node u lies to the target t the higher its $CBC(u;t)$ index.



a. MAX search pattern b. MAX search pattern

Fig. 10. a) Performance of local-information-based search scheme in terms of zero- and one-hop overlap with the CBC-determined final locations. b) One-hop overlap between the egoCBC-/CBC-determined final locations and the targets *i.e.*, the hit-rate of the search scheme.

V. CONCLUSIONS

The paper has questioned to what extent the original centrality metrics can be substituted by their computationally friendly local approximations in router-level topologies. First, the metrics are shown to exhibit high rank-correlation with their local counterparts across all datasets (22 ISPs) but one. On the other hand, the match between the two variants is much worse when we compare the top-k nodes selected by each of them. Then, we tried to assess what the algebraic values of the correlation coefficients reveal regarding the performance of local centrality-driven network functions. Both a simple navigation and a search scheme employing local centrality metrics produce significantly different navigation patterns and lower hit-rates, respectively, than their counterparts with the original global metrics. These results suggest that, despite the positive correlations, local variants can hardly offer effective approximations to the original metrics. Our work essentially

warns against relying on the correlation indices for assessing the match between ego- and sociocentered variants of centrality metrics.

REFERENCES

[1] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge Univ Pr, 1994.

[2] E. M. Daly and M. Haahr, "Social network analysis for information flow in disconnected delay-tolerant manets," *IEEE Trans. Mob. Comput.*, vol. 8, no. 5, pp. 606–621, 2009.

[3] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay-tolerant networks," *IEEE Trans. Mob. Comput.*, vol. 10, no. 11, pp. 1576 –1589, nov. 2011.

[4] W. K. Chai, D. He, I. Psaras, and G. Pavlou, "Cache "less for more" in information-centric networks," in *Proc. of the 11th IFIP Networking*, Prague, Czech Republic, May 2012.

[5] L. C. Freeman, "Centered graphs and the structure of ego networks," *Mathematical Social Sciences*, no. 3, pp. 291–234, 1982.

[6] A. Vázquez *et al.*, "Large-scale topological and dynamical properties of the internet," *Phys. Rev. E*, vol. 65, no. 6, p. 066130, Jun 2002.

[7] L. A. Adamic *et al.*, "Search in power-law networks," *Physical Review E*, vol. 64, no. 4, Sep. 2001.

[8] P. Pantazopoulos *et al.*, "Efficient social-aware content placement for opportunistic networks," in *IFIP/IEEE WONS*, Slovenia, 2010.

[9] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, vol. 25, pp. 163–177, 2001.

[10] U. Brandes and C. Pich, "Centrality Estimation in Large Networks," *Int'l Journal of Birfucation and Chaos,*, vol. 17(7), pp. 2303–2318, 2007.

[11] R. Geisberger, P. Sanders, and D. Schultes, "Better approximation of betweenness centrality." in *SIAM ALENEX*, 2008, pp. 90–100.

[12] S. Borgatti and M. Everett, "A Graph-theoretic perspective on centrality," *Social Networks*, vol. 28, no. 4, pp. 466–484, Oct. 2006.

[13] M. J. Newman, "A measure of betweenness centrality based on random walks," *Social Networks*, vol. 27, no. 1, pp. 39 – 54, 2005.

[14] Y. sup Lim *et al.*, "Online estimating the k central nodes of a network," in *IEEE Network Science Workshop (NSW'11)*, june 2011, pp. 118 –122.

[15] A.-M. Kermarrec *et al.*, "Second order centrality: Distributed assessment of nodes criticity in complex networks," *Computer Communications*, vol. 34, no. 5, pp. 619 – 628, 2011.

[16] S. Nanda and D. Kotz, "Localized bridging centrality for distributed network analysis," in *IEEE ICCCN '08*, Virgin Islands, August 2008.

[17] K. Wehmuth and A. Ziviani, "Distributed assessment of the closeness centrality ranking in complex networks," in *SIMPLEX*, NY, USA, 2012.

[18] K.-I. Goh *et al.*, "Universal behavior of load distribution in scale-free networks," *Phys. Rev. Lett.*, vol. 87, no. 27, Dec 2001.

[19] P. Marsden, "Egocentric and sociocentric measures of network centrality," *Social Networks*, vol. 24, no. 4, pp. 407–422, October 2002.

[20] M. Everett and S. P. Borgatti, "Ego network betweenness," *Social Networks*, vol. 27, no. 1, pp. 31–38, 2005.

[21] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[22] P. Pantazopoulos, M. Karaliopoulos, and I. Stavrakakis, "Centrality-driven scalable service migration," in *23rd International Teletraffic Congress (ITC'11)*, San Francisco, USA, Sept. 2011.

[23] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[24] J.-J. Pansiot, "mrinfo dataset," : http://svnet.u-strasbg.fr/mrinfo/.

[25] J.-J. Pansiot *et al.*, "Extracting intra-domain topology from mrinfo probing," in *Proc. PAM*, Zurich, Switzerland, April 2010.

[26] N. T. Spring *et al.*, "Measuring ISP topologies with rocketfuel." *IEEE/ACM Trans. Netw.*, vol. 12, no. 1, pp. 2–16, 2004.

[27] Rocketfuel: An ISP Topology Mapping Engine. [Online]. Available: http://www.cs.washington.edu/research/networking/rocketfuel/

[28] The CAIDA UCSD Macroscopic Internet Topology Data Kit (ITDK) - [ITDK 2011-10]. [Online]. Available: http://www.caida.org/data/active/internet-topology-data-kit/

[29] T. Britton *et al.*, "Generating simple random graphs with prescribed degree distribution," *J. STAT. PHYS*, vol. 124, no. 6, 2005.

[30] A.-L. Barabâsi and E. Bonabeau, "Scale-free networks," *Scientific American*, vol. 288, no. 60-69, 2003.

[31] H. P. Thadakamalla, R. Albert, and S. R. Kumara, "Local search and heterogeneities in weighted complex networks," in *NetSci'06*, Bloomington, IN, USA, May 2006.