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

Panagiotis Pantazopoulos, Merkourios Karaliopoulos, and Ioannis Stavrakakis

Department of Informatics and Telecom., National & Kapodistrian University of Athens,
Ilissia, 157 84 Athens, Greece
{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.

1 Introduction

Computer networks are generally complex systems, typically owing their complexity to their topological structure and its dynamic evolution but also the heterogeneity of their nodes. *Social Network Analysis* (SNA) [21] has been lately viewed as a solid analytical framework for understanding their structural properties and the interactions of their elements. The expectation within the networking community is that socio-analytical insights can benefit the analysis of network structure and the design of efficient protocols. Indeed, there has been evidence that such insights can improve network functions such

as content-caching strategies in wired networks [5] and routing/forwarding in opportunistic networks [6].

Common denominator to these efforts is the use of SNA-driven metrics for assessing the relative centrality (*i.e.*, importance) 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 self-organized environments lacking centralized management operations, in particular, it may not even be an option at all. A more realistic alternative for assessing node centrality draws on its ego network, *i.e.*, the subgraph involving itself, its 1-hop neighbors, and their interconnections. *Egocentric* measurements, carried out within their immediate locality, let nodes derive *local approximations* of their centrality.

Lending to simpler computations, egocentric[1] metrics have, in fact, found their way into protocol implementations [6], [5]. Nevertheless, the capacity of these local approximations to substitute the globally computed sociocentric metrics is almost always taken for granted rather than evaluated. Over Internet router-level topologies, in particular, a systematic study of the approximative power of egocentric metrics is missing.

In this paper, we focus on node centrality metrics, which are the most commonly used in networking protocols. Besides the well known Betweenness Centrality (BC) metric, we consider the Conditional BC (CBC), the destination-aware BC variant proposed in [16], which is particularly suited to many-to-one communication and data flow typologies. We first question how well do sociocentric metrics, computed under global topological information, *correlate* with their egocentric variants, as computed locally over the nodes' ego networks. Since in most protocol implementations, it is the *ranking* of the metric values that matters rather than their *absolute values*, we measure this correlation over the node rankings they produce, via their full *rank*-correlation and the overlap in the top-*k* node sets in the two rankings. Then, contrary to previous works, we proceed to study what the measured correlation coefficients can reveal regarding the capacity of rank-preserving local centrality metrics to substitute the original global metrics in protocol primitives.

Related work: Computing centrality values typically requires global topological information. BC computations, in particular, can be carried out by exact yet efficient [3] or approximation algorithms [4]. The latter seek to provide accurate sampling methods over the considered network node pairs. Some recently proposed random walk schemes that reveal high BC [11] or central nodes in a general sense [10], provide *distributed*-fashion approaches but require gathering beyond-local-scope information. The efforts to devise *local* approximations of the global centrality metrics span various disciplines and usually study *whether* do the former correlate with the latter. From a networking standpoint, local approximations of centrality metric have been proposed for the bridging [13] and closeness centrality [22]. Both works report high positive rank-correlation scores between the two counterparts over synthetic and real world networks, but do not explore whether this is sufficient to benefit networking operations. Betweenness Centrality, on the other hand, is locally approximated over real-world network topologies

---

[1] The terms egocentric and local as well as sociocentric and global are used interchangeably in the paper.

with the node degree, or Degree Centrality (DC), due to the evidenced linear correlation between the mean BC and DC metrics [20]. However, this holds for AS-level Internet maps that can exhibit different connectivity properties compared to the router-level topologies; or, in the case of *load*, a BC variant, for scale-free networks [9]. Finally, some social studies provide experimental evidence for positive correlation between sociocentric and egocentric BC over *small social* and *synthetic* networks [12], [7].

Our contribution: We report on the BC *vs.* egoBC, BC *vs.* DC, CBC *vs.* egoCBC *correlation* over synthetic topologies and then focus on real-world router-level topologies of sizes up to 80K nodes, coming from 20 different ISPs (Section 3). Our synthetic networks yield significant rank-correlation that weakens as their size increases. In router-level topologies, in almost all cases, we measure high rank correlation (0.8-0.9), which becomes even higher when we compute the egocentric betweenness variants over generalized ego networks, corresponding to their 2-hop neighborhoods. In each case, we analyze the time complexity and message overhead related to the computation of the metrics. (Section 2).

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 4) that high rank-correlation values between the two BC variants are poor predictors for the capacity of the local approximations to substitute the original global metrics in actual network protocol operations.

2 Sociocentric *vs.* egocentric centrality metrics

The sociocentric metrics of betweenness centrality (BC) and the conditional betweenness centrality (CBC), along with their egocentric counterparts[2] are first presented. Then, the complexity savings achieved by the latter, are discussed.

2.1 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 [8]. 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 [17] 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.

---

[2] The degree centrality (DC) is by default an egocentric centrality metric.
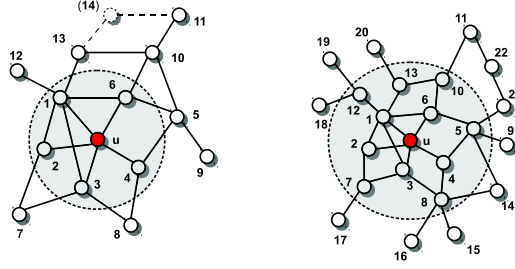
a. Ego network of node u ($r = 1$).   b. Ego network of node u ($r = 2$).

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 $egoCBC(u; 11, 2)$.

2.2 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 rth-order ego network as follows. Let N_r^u be the set of nodes that form the r-hop neighborhood 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 rth-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 $|E_1^u| = DC(u) + CC(u) \cdot \binom{DC(u)}{2}$ edges, where $CC(u)$ is the clustering coefficient [14] 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 : $egoBC(u; r) = BC(u)_{|V = V_r^u}$. 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
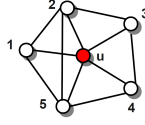
Fig. 2. Toy-example computation: $egoBC(u;1) = 10(1 - \frac{6}{10}) - [2(1 - \frac{1}{2}) + 2(1 - \frac{1}{3})] = \frac{5}{3}$

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 have:

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

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). 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.

2.3 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 [3], 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 [3].

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 [7]. The computation involves a multiplication of an $O(d_{max})$-size square matrix and trivial condition checks. For $r = 2$, we can employ [3] replacing $|V|$ with d_{max}^2. Finally, DC is considered to be immediately available to every node.

Table 1. Complexity comparison of socio- *vs.*ego-centric metrics

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

---

[3] The algorithm in [3] 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 $u \in V$.

3 Experimental correlation-operations study between socio- and egocentric centrality metrics

3.1 Correlation coefficients and network topologies

In comparing the ego- with sociocentric metrics, we are concerned with (rank) correlation-operations. Protocol implementations that utilize highly central nodes, usually care more about the way the metric *ranks* the network nodes rather than their absolute metric values; this is how a degree-based search scheme explores unstructured P2P networks [2] or egoBC is used in DTN forwarding [6]. Alternatively, we may need to employ only a subset of k nodes with the top centrality values [17] and therefore, we also 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 ρ, which assesses how monotonic is the relationship between the ranks of the two centrality variables. For the sake of completeness we also measure the well-known Pearson correlation (see [18] for the relevant formulas). The r_{Prs} coefficient also 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] and mark a significant correlation between the considered metrics when they are close to 1. For the target-node-dependent CBC values, we present the correlation averages along with the 95% confidence intervals estimated over an at least 7% of the total locations, sample.

Regarding the networks we experiment with, our emphasis is on the real-world intradomain Internet topologies. Nevertheless, synthetic graphs with distinct structural properties such as the rectangular grid have been also employed to provide insights. The router-level ISP topologies do not have the predictable structure of the synthetic ones and typically size *up to* a few thousand nodes. Yet, over these topologies networking protocols designed to cope with self-organization requirements, will seek to utilize local centrality metrics [17]. We have experimented with three sets of router-level topologies:

mrinfo topologies: The dataset we consider includes topology data from 850 distinct snapshots of 14 different AS topologies, corresponding to Tier-1, Transit and Stub ISPs [15]. 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 experiment with a representative subset of Tier-1 and Transit topologies.

Rocketfuel topologies: The Rocketfuel technique [19] 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 includes measurements from 800 vantage points serving as `traceroute` sources. Innovative techniques such as BGP directed probing and IP identifiers, have been applied to reduce the number of probes, and discover the different interface IP addresses that belong to the same router (*i.e.*, alias resolution), respectively.

CAIDA topologies: The most recent of our datasets [1] was collected during Oct-Nov 2011 by CAIDA performing `traceroute` probes to randomly-chosen destinations from 54 monitors worldwide. Parsing the provided separate file that heuristically assigns an AS to each node found, we have determined the router-to-AS ownership and

Table 2. Correlation study between BC and egoBC on grid networks

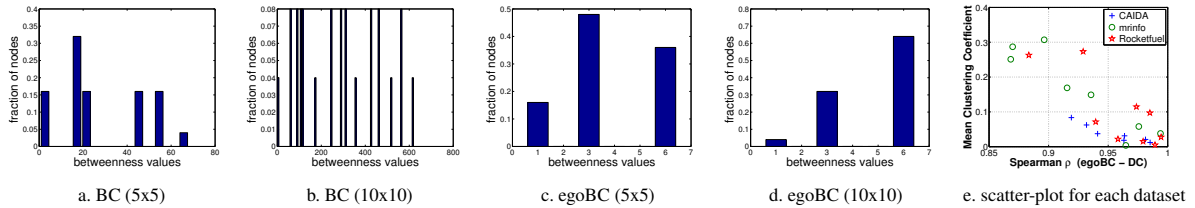| Grid size	Diameter / Mean degree	Spearman ρ	
		ego-network **(r=1)**	ego-network **(r=2)**
5x5	8 / 3.200	0.9195	0.9679
10x10	18 / 3.600	0.8400	0.9556
20x20	38 / 3.800	0.6802	0.8459
50x50	98 / 3.920	0.2429	0.2942
60x8	66 / 3.717	0.5735	0.6336
90x8	96 / 3.728	0.5390	0.5870
150x8	156 / 3.737	0.4584	0.4181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

a. BC (5x5)   b. BC (10x10)   c. egoBC (5x5)   d. egoBC (10x10)   e. scatter-plot for each dataset

Fig. 3. a)-d) Probability distribution of BC and egoBC values for scaling size of a grid network. e) egoBC-DC correlation scales linearly with <CC> values.

subsequently have extracted out of the raw data files the topologies of the nodes operated by certain ASes. Our effort was to discover the largest ISP networks in the dataset. With all three datasets we avail a rich experimentation basis of a diverse Internet topologies' set that can minimize the effect of measurement technique errors.

3.2 Experimental results

We choose to spend more effort experimenting on the BC-egoBC, BC-DC correlation debates, expected to attract more interest (see Section 4) than the limited-scope of CBC.

egoBC vs.BC : In grid topologies, ego networks have fixed size depending on their position, *i.e.*, corner, side, or internal nodes. Thus, the egoBC index may only exhibit a small number of different values (*e.g.*, 1, 3 and 6 with respect to the node's location when $r = 1$). When grid dimensions grow larger, the number of shortest paths between any node pair grows exponentially, resulting in a richer spectrum of BC values over the grid nodes. On the other hand, the possible egoBC values remain the same; only the distribution of grid nodes over these values changes (see Fig. 3a-d). Consequently, in Table 2, ρ values decrease monotonically with the grid size. Same behavior and reasoning also holds for the line network.

Our findings for the real-world ISP topologies are listed in Table 3. Even with measurements within the first-order ego network, there is high positive rank correlation between BC and egoBC. On the other hand, the Pearson coefficient suggests looser yet positive association. When egoBC is computed in the second-order ego network, both correlation coefficients become even higher and also more similar with each other. The structural characteristics of the considered ISP topologies differ from the grid topologies; their diameter and clustering coefficients 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 that favors high correlation. A notable exception is the

Table 3. Correlation study between BC-egoBC and BC-DC on router-level ISP topologies

DataSet	ID	ISP(AS number)	$<\mathcal{CC}>$	Diameter	Size	$<degree>$	BC *vs.*ego-BC				BC *vs.*DC	
							Spearman ρ		Pearson r_{Prs}		Spearman ρ	Pearson r_{Prs}
							ego-net. **r=1**	ego-net. **r=2**	ego-net. **r=1**	ego-net. **r=2**		
	35	Global Crossing(3549)	0.479	9	100	3.78	0.9690	0.9853	0.7029	0.9255	0.8506	0.6714
m	33	NTTC-Gin(2914)	0.307	11	180	3.53	0.9209	0.9565	0.7479	0.8561	0.8180	0.6664
r	13	Level-3(3356)	0.169	25	378	4.49	0.2708	0.9393	-0.0918	0.7982	0.1953	-0.0813
i	12	-//-	0.149	28	436	4.98	0.2055	0.9381	-0.1217	0.7392	0.1696	-0.1128
n	20	Sprint(1239)	0.287	16	528	3.13	0.9866	0.9928	0.5805	0.8488	0.8543	0.6815
f	38	Iunet(1267)	0.231	12	645	3.75	0.8790	0.9516	0.9094	0.9568	0.8549	0.7708
o	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	61	Ebone(1755)	0.115	13	295	3.68	0.9736	0.9860	0.6856	0.8895	0.9443	0.7457
O	62	Tiscali(3257)	0.028	14	411	3.18	0.9522	0.9659	0.6073	0.9281	0.9464	0.7103
C	63	Exodus(3967)	0.273	14	353	4.65	0.9125	0.9792	0.6100	0.9061	0.8204	0.6241
K	64	Telstra (1221)	0.015	15	2515	2.42	0.9990	0.9990	0.3336	0.7565	0.9783	0.5172
E	65	Sprint(1239)	0.022	13	7303	2.71	0.9980	0.9990	0.4770	0.7977	0.9562	0.6537
T	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	74	ChinaTelecom(4134)	0.083	19	81121	3.97	0.8324	0.8986	0.7861	0.9714	0.7370	0.8795
D	75	FUSE-NET(6181)	0.018	10	1831	2.38	0.9903	0.9763	0.6205	0.8574	0.9536	0.7445
A	76	JanetUK(786)	0.031	24	2259	2.26	0.9819	0.9834	0.4444	0.8506	0.9450	0.5765

`mrinfo` Level-3 ISP topology (IDs 12,13) which we comprehensively study (and visualize) in [18]. Herein, we provide only a brief explanation. Datasets 12,13 were found to avail some clustered structures of nodes that, interestingly, exhibit higher egoBC than global BC values; it is these nodes that distort the desired linear relation between the centrality counterparts. Finally, note that Table 3 results imply that the localized bridging centrality [13], which uses egoBC to approximate BC, is also highly correlated with its global counterpart in these topologies.

DC vs.BC: In grid topologies we have observed the same correlation degradation with the network size, explained along the aforementioned BC-spectrum arguments. In all router-level topologies (Table 3) we find high Pearson and even higher Spearman correlation although consistently lower than the corresponding egoBC *vs.*BC one, at least for the Spearman ρ. As such, the previously reported [20] DC-BC correlation over AS-level topologies is extended, by our results, to the router-level ones. Finally, the earlier observed egoBC-BC rank correlation can be further justified on the grounds of the high DC-BC Spearman values; the router-level topologies exhibit vanishing clustering coefficients and thus, the egoBC metric attains similar values to DC (see Eq. 1). This is depicted in Fig. 3.e where the egoBC-DC correlation scales linearly with the mean clustering coefficients, especially in the CAIDA set.

egoCBC vs.CBC : We now assess the ego variant of the CBC metric (r=1). Table 4 suggests significant positive rank correlation in all studied ISP topologies, even for the outlier case of the `mrinfo` Level-3 networks. Intuitively, the correlation between CBC counterparts is expected to be higher than the one between BC counterparts; by neglecting the world outside of the ego network, the egoBC inaccuracies (compared to the globally determined BC) may arise anywhere across the network. On the contrary, $egoCBC(u; t, r)$ 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 : Finally, we assess the extent to which the nodes with the top-k values of local centrality metrics coincide with the corresponding ones determined by the global BC. Interestingly, Table 5 shows that this overlap is low, at least for the first-order ego network measurements. This does not actually contradict our previous results since correlation is 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 [7]. Along this thread, our comprehensive study [18] has shown that the actual association between the two metric variants is not determined solely by the degree distribution. Finally, the low top-k overlap scores serve as a warning sign as to what the high coefficients can reveal about the practical implications of local centrality metrics.

ID	35	33	13	12	20	44
Spearman ρ	0.9489	0.9554	0.7336	0.7035	0.9847	0.9902
Conf. Interv.	0.013	0.003	0.007	0.005	0.003	0.001
ID	50	61	62	63	75	76
Spearman ρ	0.9739	0.8423	0.9321	0.7641	0.9961	0.9853
Conf. Interv.	0.009	0.027	0.016	0.023	0.005	0.002

Table 4. Rank-correlation between CBC and ego-CBC (r=1).

	$k=10$			$k=30$		
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 5. Overlap(%) between nodes with the top-k local centrality and BC values.

4 Practical utility of local centrality metrics

We now study basic network operations where local centrality metrics are employed, encouraged by the high corresponding correlation, to substitute the global ones. Our aim is to assess whether such options can actually be effective in networking practice.

A local-centrality-driven navigation scheme: In unstructured power-law P2P networks, it has been shown [2] that a high-degree (file) 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 great number of links pointing to high degree nodes. 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 refers to the destinations' exact matching. Regarding the final locations, figs. 4a-d) show that the local metrics ($r=1$) can hardly be effective aliases for the global ones. The crawler, driven by local metrics, is measured [18] to consistently take from 0.4 up to
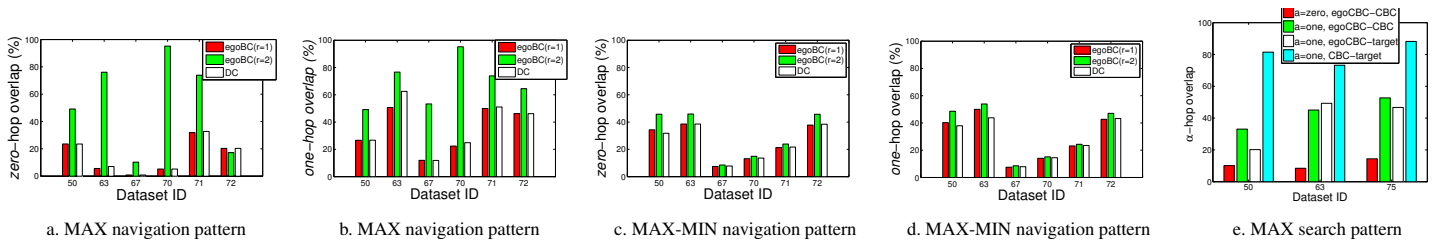
| a. MAX navigation pattern | b. MAX navigation pattern | c. MAX-MIN navigation pattern | d. MAX-MIN navigation pattern | e. MAX search pattern |

Fig. 4. a-e) Overlap between the final locations achieved with local and global centrality as driver; in e) the two rightmost bars of each dataset show the egoCBC/CBC-driven search hit-rates.

2.3 on average less hops than when it is BC-driven, failing to identify the same navigation pattern *i.e.*, sequence of central nodes. In view of the high correlation between the involved metrics, we have tried to shed some light on the somewhat counterintuitive, poor navigation performance [18]. Removing sequentially four nodes from a toy topology of initially perfect rank correlation (*i.e.*, $\rho = 1$) between the two counterparts, we measured it reducing to 0.9953. At the same time the $zero$-hop overlap for the MAX pattern drastically diminished from 100% to 61.90%. Clearly, the numerical summaries the coefficients provide, fail to capture in micro-level the relative significance of each node which determines the scheme's performance.

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. 4.e 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 [2], figure 4.e (two rightmost bars) 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 [2], 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 overlap values when the global metric drives the MAX search pattern, since the closer node u lies to the target t the higher its $CBC(u; t)$.

5 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 (20 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 network functions, when the original metrics are substituted by their local approximations. 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 substitutability of ego- and sociocentered variants of centrality metrics.

References

[1] The CAIDA UCSD Macroscopic Internet Topology Data Kit-[ITDK 2011-10], `http://www.caida.org/data/active/internet-topology-data-kit/`

[2] Adamic, L.A., et al.: Search in power-law networks. Physical Review E 64(4) (Sep 2001)

[3] Brandes, U.: A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25, 163–177 (2001)

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

[5] Chai, W.K., et al.: Cache "less for more" in information-centric networks. In: Proc. of the 11th IFIP Networking. Czech Republic (2012)

[6] Daly, E.M., Haahr, M.: Social network analysis for information flow in disconnected delay-tolerant manets. IEEE Trans. Mob. Comput. 8(5), 606–621 (2009)

[7] Everett, M., Borgatti, S.P.: Ego network betweenness. Social Networks 27(1), 31–38 (2005)

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

[9] Goh, K.I., et al.: Universal behavior of load distribution in scale-free networks. Phys. Rev. Lett. 87(27) (Dec 2001)

[10] Kermarrec, A.M., et al.: Second order centrality: Distributed assessment of nodes criticity in complex networks. Comp. Com. 34 (2011)

[11] Lim, Y., et al.: Online estimating the k central nodes of a network. In: IEEE Network Science Workshop (NSW'11). pp. 118 –122 (June 2011)

[12] Marsden, P.: Egocentric and sociocentric measures of network centrality. Social Networks 24(4), 407–422 (October 2002)

[13] Nanda, S., Kotz, D.: Localized bridging centrality for distributed network analysis. In: IEEE ICCCN '08. Virgin Islands (August 2008)

[14] Newman, M.E.J.: The Structure and Function of Complex Networks. SIAM Review 45(2), 167–256 (2003)

[15] Pansiot, J.J., et al.: Extracting intra-domain topology from mrinfo probing. In: Proc. PAM. Zurich, Switzerland (April 2010)

[16] Pantazopoulos, P., et al.: Efficient social-aware content placement for opportunistic networks. In: IFIP/IEEE WONS. Slovenia (2010)

[17] Pantazopoulos, P., et al.: Centrality-driven scalable service migration. In: 23rd Int'l Tele-traffic Congress (ITC'11). San Francisco, USA (2011)

[18] Pantazopoulos, P., et al.: On the local approximations of node centrality in Internet router-level topologies. Tech. rep. (January 2013), `http://anr.di.uoa.gr/index.php/publications`

[19] Spring, N.T., et al.: Measuring ISP topologies with rocketfuel. IEEE/ACM Trans. Netw. 12(1), 2–16 (2004)

[20] Vázquez, A., et al.: Large-scale topological and dynamical properties of the Internet. Phys. Rev. E 65(6), 066130 (Jun 2002)

[21] Wasserman, S., Faust, K.: Social network analysis. Cambridge Univ Pr (1994)

[22] Wehmuth, K., Ziviani, A.: Distributed assessment of the closeness centrality ranking in complex networks. In: SIMPLEX. NY, USA (2012)