

Assessing the value of localized centrality metrics

Panagiotis Pantazopoulos Merkourios Karaliopoulos and Ioannis Stavrakakis

Department of Informatics and Telecommunications

National & Kapodistrian University of Athens

Ilissia, 157 84 Athens, Greece

Email: {ppantaz, mkaralio, ioannis}@di.uoa.gr

Abstract

Acquiring the full global information is impractical, if feasible at all, in many networks with distributed operation and self-organization features. To meet scalability requirements practical protocol implementations could use local information instead, drawn from the nodes' *ego-networks*, the Social Network Analysis (SNA) counterpart of *centered graphs*. However, in almost all these efforts the capacity of local *ego-centered measurements* to substitute global information is not evaluated but rather taken for granted. It is this assumption that we effectively put under the microscope.

We focus on network centrality metrics and explore to what extent *sociocentric* metrics determined under complete network-wide information (rank-)correlate with their ego network counterparts. We generalize the definition of ego network, to comprise the ego's r -hop neighborhood, respectively, and analyze the involved accuracy vs. complexity tradeoff. The correlation of the global and ego metric variants is found high in synthetic and even higher in real-world network topologies of several hundred nodes size. We discuss the impact of network topology on this correlation and possible network functions that can benefit from rank-preserving ego-centered measurements. Our findings can serve as evidence for the practical feasibility of "socioaware" networking.

I. INTRODUCTION

Recent approaches to data networking draw heavily on the communication patterns established by the end-users. Human relationships and their inherent properties can serve as a basis for developing a user centric communication paradigm, where the basic networking functionality is built around the end-user [1], [2]. A solid theoretical framework for processing social information and analyzing social structures underlying these relationships is provided by *Social Network Analysis* (SNA) [3]. SNA has, thus, generated a lot of interest within the networking community and expectations that its insights could benefit the design of more efficient network protocols.

Indeed, there is growing evidence that SNA insights and techniques can improve the design of such functions as routing in opportunistic networks [4] [5], and service facility placement in wired networks [6]. Common denominator

to all these efforts is the use of some SNA-driven metric 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 network environments it may not even be an option at all as for example is the case in emerging self-organizing networking paradigms lacking centralized network management processes. A more realistic alternative for assessing node centrality may be based on its ego network (SNA term) or centered graph (graph-theoretic term) [7], *i.e.*, the subgraph involving itself, its 1-hop neighbors, and their interconnections. Nodes can acquire a local estimate of their centrality through *egocentric measurements* in their immediate locality. The computation of egocentric metrics is apparently less complex and, in fact, part of practical protocol implementations [4] [5]. Nevertheless, the capacity of these local egocentric metrics to substitute the globally computed sociocentric metrics is almost never evaluated but rather taken for granted. It is this assumption that our paper questions.

We set our focus on network 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 wireless sensor networks and, more generally, networks with many-to-one data flow typologies. An important remark we make is that in several cases, it is the *order* of the metric values that matters rather than their *absolute values*. Therefore, we investigate whether and how much the sociocentric metrics, computed under global topological information, *rank-correlate* with their egocentric counterparts, computed locally over the nodes' ego networks. In a further step, we carry out the correlation analysis over a generalized definition of the ego network including the 2-hop neighbors of the ego node. While analyzing the corresponding time complexity and message overhead for each alternative, we derive a full view of the complexity *vs.* accuracy tradeoff.

Our evaluation is carried out over both synthetic graphs and real world topologies up to 1000 nodes. For relative small sizes, the former networks yield significant correlation (between the two centrality variants) that weakens as their size increases. For the latter networks we measure high correlation values in the order of [0.8-0.9] in almost all cases for both centrality metrics. Conceivably, we obtain better results when the 2-hop neighbors are concluded in the ego network.

The paper is structured as follows: we first elaborate on the socio- and ego-centric centrality variants, compute the latter over 1- and 2-hop ego networks, and compare their computational complexity in Section II. Then in Section III we present the results of our correlation study and outline example applications of egocentric metrics. We discuss the relevant literature in Section IV and conclude the paper in V.

II. SOCIOCENTRIC *vs.* EGOCENTRIC CENTRALITY METRICS

We briefly present the two sociocentric centrality metrics our study addresses, the betweenness centrality (BC) and the conditional betweenness centrality (CBC), and their egocentric counterparts¹. We discuss the complexity

¹This comparison is trivial for the popular degree and closeness centrality metrics. The former is effectively an egocentric metric itself; the latter turns out to be uninformative under an egocentric view since all shortest paths from an individual node to its first-order neighbours have length one [9].

savings of locally measured metrics in section II-C and devote section III to assessing how well they can substitute the sociocentric ones.

A. Globally computed centrality indices

Consider an arbitrary node pair (s, t) over a connected 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 , then the betweenness centrality of node u , $BC(u)$, equals

$$BC(u) = \sum_{\substack{s, t \in V \\ s \neq t \neq u}} \frac{\sigma_{st}(u)}{\sigma_{st}} \quad (1)$$

Effectively, $BC(u)$ assesses the importance of a network node for serving information that flows over shortest paths in the network [9]. Whereas $BC(u)$ is an average over all network node pairs, the second centrality metric we consider, the conditional betweenness centrality index (CBC), captures the topological centrality of node u with respect to a *specific* destination node t [6] and is given by:

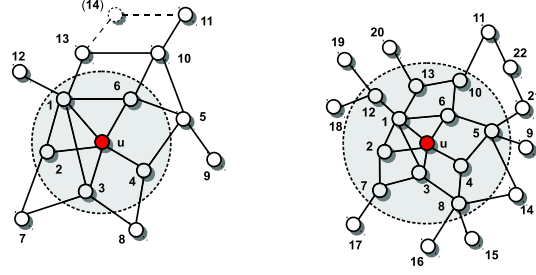
$$CBC(u; t) = \sum_{\substack{s \in V \\ s \neq t, u \neq t}} \frac{\sigma_{st}(u)}{\sigma_{st}} \quad (2)$$

with $\sigma_{st}(s) = 0$. Therefore, CBC is particularly suited to settings where information flows are directed towards a particular node with discrete network functionality. Sink nodes collecting measurement data in wireless sensor networks and gateways routing traffic at the border between (wireless) access networks and the core network are two examples of such settings.

B. Locally computed centrality metrics - ego networks

The computation of the two centrality metrics inline with their sociocentric definition in (1) and (2) requires information about the whole network topology and implies good coordination across the network as well as computational and message load overheads. In distributed network settings, where nodes are energy constrained and no explicit centralized network management processes are in place, these computations are not favorable or not an option at all. The apparent choice to approximate the quantities of eq. 1 and 2 drawing on local information appears to be the use of degree (centrality). Indeed, the degree of a node has been already reported to linearly relate to its BC value at the autonomous system level (AS maps) of Internet topologies [10]. Moreover, in scale-free networks a node's degree has been found under conditions to be proportional to its *load* *i.e.*, an equivalent metric to BC [11]. Nevertheless, in the case of typical shortest-path routing approaches the interpretation of degree centrality seems problematic; it cannot be clearly considered as a measure of the network flow amount that a given node controls. It rather suits for assessing a random-walk based information dissemination process across the network [12]. Accordingly, the degree centrality is not capable of capturing the destination-awareness notion as in the case of CBC.

Techniques to approximate the centrality metrics through localized variants can be borrowed from the social network analysis and the concepts of ego networks and ego-centered measurements. In the so-called *ego-network*



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 $egoCBC(u; 11, 2)$.

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. The counterpart of ego network in graph-theoretic terms (the shaded area in Fig. 1) is called centered graph [7] and includes a given node u together with its 1-hop neighbor nodes.

Hereafter, we generalize the definition of the ego-network 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 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 r^{th} -order ego network of node u is the centered graph $G_r^u = (V_r^u, E_r^u)$, where the set of nodes and edges are $V_r^u = \{N_r^u, u\}$ and $E_r^u = \{(i, j) \in E : i, j \in N_r^u\}$, respectively. For $r = 1$ the network G_1^u corresponds to the original ego network definition and consists of $|V_1^u| = d(u) + 1$ nodes and $|E_1^u| = d(u) + CC(u) \cdot \binom{d(u)}{2}$ edges, where $d(i)$ is the degree and $CC(i)$ the clustering coefficient of node i . Practically, values of $r > 2$ would tend to cancel the advantages that local ego-centered measurements induce.

Accordingly, the betweenness 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 (1) over the graph G_r^u

$$egoBC(u; r) = BC(u)|_{V=V_r^u}. \quad (3)$$

The value that $egoBC(u, r)$ attains can be significantly different than the degree value of node u and thus, we can not rely on the degree-betweenness reported correlation [10] to claim association between local and global BC values. To further illustrate the relation between the localized betweenness and other well-known network properties we elaborate on the computation formula of $egoBC(u, 1)$. In the figure 2.a the ego-node u is connected to five first-neighbor nodes, initially not interconnected with each other. The number of (2-hops-long) paths traversing u is equal to the different pairs of neighbors i.e., the the combination of 5 taken 2. In case the dotted links 1-2 and 3-4 exist, we need to subtract the number of 2-hops paths that are no more used from corresponding pairs since the nodes involved can now communicate directly. Clearly, these paths are as many as the direct paths linking the first-neighbor nodes one another. The requested quantity is the numerator of the clustering coefficient ($CC(u)$)

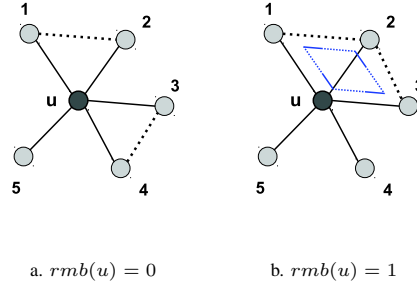


Fig. 2. Computing ego betweenness for an example ego-network.

since the latter is defined as the ratio between the number of connections among the first neighbors of u and its maximum possible value. Finally, when the configuration of the direct links between the first-neighbors (Fig. 2.b) leads to the formation of one or more rhombus-like loops (*i.e.*, 1-2-3- u -1) including the ego plus 3 neighboring nodes, only one (1- u -3) out of the two shortest paths between the first-neighbors passes through u . Summing up, for the general case of $d(u)$ first-neighbors of node u and $rmb(u)$ rhombus-like loops around u , we have:

$$egoBC(u; 1) = \begin{cases} \binom{d(u)}{2}(1 - CC(u)) - rmb(u)/2 & \text{if } d(u) > 1 \\ 0 & \text{if } d(u) = 1 \end{cases} \quad (4)$$

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 $e_r(u; t) = \{t_e \in N_r^u : h(u, t') + h(t', t) = h(u, t)\}$, *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 and $e_2(u; 11) = \{10\}$, $e_2(u; 14) = \{5, 8\}$ for the G_2^u . 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

$$egoCBC(u; t, r) = \sum_{\substack{s \in V_r^u \\ t' \in e_r(u; t)}} \frac{\sigma_{st'}(u)}{\sigma_{st'}} \mathbf{1}_{\{h(s, t') \leq h(s, l), l \in e_r(u; t)\}} \quad (5)$$

Again, in Fig. 1a, node 4 contributes to the egoCBC value of node u since its shortest path to the single exit node 6 passes through u , although it has a shorter path of length three 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 the 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 sociocentric and egocentric measurements in (1)-(5) are valid under weighted and unweighted graphs, we focus on the latter ones (note that (4) is a formula for the $egoBC(u, 1)$ computation

in unweighted graphs). The way network link weights affect the correlation of the two types of metrics is clearly worth of a separate study.

C. Complexity comparison between socio- and egocentric centrality metrics

We discuss briefly 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 the computation algorithm.

Sociocentric centrality metric computation. 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 network node can compute the BC metric values of all other nodes in the network. An efficient way to do this is to invoke Brandes' algorithm in [13], featuring $O(|V| \cdot |E|)$ complexity for unweighted graphs and $O((|E| + |V|) \cdot |V| \log |V|)$ 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 computation².

Egocentric centrality metric computation. 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 egocentric computation of egoBC and egoCBC for $r = 1$ can be carried out as in [14]. The computation involves a multiplication of an $O(d_{max})$ -size square matrix and trivial condition checks. For $r = 2$, we can compute the two metrics with Brandes' algorithm, replacing $|V|$ with d_{max}^2 .

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

²The algorithm in [13] 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 $u \in V$.

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)	$O(d_{max}^3)$	$O(2 \cdot E)$
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)	$O(d_{max}^4)$	$O(2 \cdot d_{max} \cdot E)$

III. EXPERIMENTAL CORRELATION STUDY OF SOCIO- vs. . EGOCENTRIC CENTRALITY METRICS

A. Correlation coefficients and network topologies

In comparing the ego- with sociocentric metrics, we are mostly concerned with their *rank* correlation. The underlying remark is that in protocol implementations drawing on social metrics, we care more about the way the metric *ranks* different nodes rather than its absolute value for each one of them. For example, this is how the egoBC metric is used in the SimBetTS [4] and BubbleRap [5] DTN routing protocols. We capture the rank correlation in the non-parametric Spearman measure of correlation, ρ , which assesses how monotonic is the relationship between the two centrality variables and is computed as follows:

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

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. The ρ value lies in [-1,1]. In case ρ is close to 1 we can only infer significant correlation between the two variables. 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. For the pairs of the socio- and ego- betweenness variants ($sB(u), eB(u)$) of each node $u \in V$, the Pearson r_{Prs} is given by the equation:

$$r_{Prs} = \frac{\sum_{u \in V} (sB(u) - \overline{sB})(eB(u) - \overline{eB})}{\sqrt{\sum_{u \in V} (sB(u) - \overline{sB})^2} \sqrt{\sum_{u \in V} (eB(u) - \overline{eB})^2}} \quad (7)$$

When the involved parameters vary along the network instances (e.g., BA graphs) or the node location (e.g., CBC values), we present the ρ and r_{Prs} averages together with the 95% confidence intervals, estimated over 20 runs.

We experiment with both synthetic and real-world topologies to cover an adequately differentiated set of physical network topologies. The synthetic graphs are of two types, Barabási-Albert (B-A) and two-dimensional rectangular grid graphs, *i.e.*, two well defined instances of graph models with very different and distinct structural properties. Real-world ISP networks, on the other hand, do not have the predictable structure and properties of the synthetic topologies and may differ substantially one from another. Yet it is over such networks that (socioaware) networking

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
60x8	66 / 3.717	0.5735	0.6336
70x8	76 / 3.721	0.5569	0.6130
80x8	86 / 3.725	0.5466	0.5978
90x8	96 / 3.728	0.5390	0.5870

protocols have to operate [6]. The dataset we consider [15] includes topology data from 850 distinct snapshots of 14 different AS topologies, corresponding to five Tier-1, five Transit and four Stub ISPs [16]. 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 such as traceroute. Herein we present and discuss results from a representative subset of the datasets (Tier-1 and Transit ISP networks) exhibiting adequate variance in size (up to 1000 nodes), diameter, and connectivity degree statistics.

B. Experimental results

We choose to spend more of our effort experimenting on the BC-egoBC correlation debate that is expected to attract more interest; BC is extensively used in numerous applications (see Section I) as opposed to the limited scope of CBC.

1) *BC vs. egoBC*: 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 the Appendix we derive analytical expressions of the egoBC computation in grid networks. 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).

For the B-A graphs³, the reported ρ coefficients are mean values computed out of 20 graph instances (with $m_0 = 5$, $m = 2$) generated according to the preferential attachment principle [17]. As shown in Table IV, the B-A topological characteristics result in better correlation between the egoBC and BC values. B-A topologies feature higher topological contrast than grid topologies, with hub nodes featuring connectivity degrees orders of size higher

³The B-A network requirement of the node degree distribution exponent being equal to 3 appears with network sizes of 1000 nodes or more. Therefore, the employed networks of a few hundreds of nodes are small to be called B-A networks; this name is only kept for the sake of clarity while typically they should be called scale-free B-A-like.

TABLE III
CORRELATION STUDY (PEARSON r_{PRS} COEFFICIENT) BETWEEN BC AND EGOBC ON BARABÁSI-ALBERT RANDOM GRAPHS

Number of nodes	Ego network $r=1$			Ego network $r=2$		
	Range of Mean clustering coeff.	Pearson r_{PRS}	95% Conf. Interval	Range of Mean clustering coeff.	Pearson r_{PRS}	95% Conf. Interval
50	[0.0948-0.3298]	0.9791	0.003	[0.1293-0.3720]	0.9936	0.0017
100	[0.0852-0.2094]	0.9758	0.005	[0.0759-0.1972]	0.9950	0.0008
200	[0.0590-0.1429]	0.9667	0.005	[0.0586-0.1787]	0.9922	0.0014
300	[0.0485-0.1053]	0.9671	0.004	[0.0449-0.1162]	0.9919	0.0012
400	[0.0259-0.0900]	0.9634	0.007	[0.0230-0.0849]	0.9910	0.0015
500	[0.0303-0.0577]	0.9690	0.005	[0.0226-0.0654]	0.9897	0.0017
600	[0.0255-0.0735]	0.9659	0.006	[0.0275-0.0557]	0.9900	0.0013

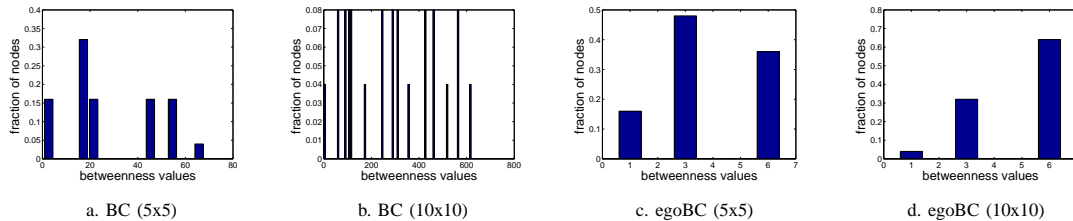


Fig. 3. Probability distribution of BC and egoBC values for scaling size of a grid network.

than other (leaf) nodes; this results in highly differentiated BC values across the network nodes. However, the egoBC values appear to exhibit equally high variance. The ego network size for $r = 1$ follows the power-law distribution of the node degree so that the percentage of B-A nodes in a single ego network may vary significantly. In the generated B-A instances, the ego network of the 200-node graph for $r = 1$ comprises up to 20.5% of the total number of nodes; whereas, the biggest ego network for $r = 2$, involves more than 62% of the nodes.

Our findings for the real-world ISP topologies are listed in table V. 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.*, Janet UK and Sprint) show clearly stronger correlation than others (*e.g.*, Telecom Italia). 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 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 both grid and B-A

TABLE IV
CORRELATION STUDY (SPEARMAN COEFFICIENT) BETWEEN BC AND EGOBC ON BARABÁSI-ALBERT RANDOM GRAPHS

Number of nodes	Ego network $r=1$			Ego network $r=2$		
	Range of Mean clustering coeff.	Spearman ρ	95% Conf. Interval	Range of Mean clustering coeff.	Spearman ρ	95% Conf. Interval
50	[0.0948-0.3298]	0.9426	0.010	[0.1293-0.3720]	0.9734	0.006
100	[0.0852-0.2094]	0.9170	0.009	[0.0759-0.1972]	0.9574	0.007
200	[0.0590-0.1429]	0.9130	0.005	[0.0586-0.1787]	0.9472	0.004
300	[0.0485-0.1053]	0.8992	0.004	[0.0449-0.1162]	0.9381	0.006
400	[0.0259-0.0900]	0.8945	0.006	[0.0230-0.0849]	0.9301	0.005
500	[0.0303-0.0577]	0.8867	0.005	[0.0226-0.0654]	0.9248	0.003
600	[0.0255-0.0735]	0.8764	0.004	[0.0275-0.0557]	0.9173	0.005

TABLE V
CORRELATION STUDY BETWEEN BC-egoBC ON INTRA-DOMAIN ISP TOPOLOGIES

DataSet	ISP (AS number)	<Clustering coeff.>	Diameter	Size	<degree>	BC vs. ego-BC				
						Spearman ρ		Pearson $r_{P_{rs}}$		
						ego-net. $r=1$	ego-net. $r=2$	ego-net. $r=1$	ego-net. $r=2$	
Tier-1	36	Global Crossing (3549)	0.546	10	76	3.71	0.9648	0.9806	0.6720	0.9197
	35	-/-	0.479	9	100	3.78	0.9690	0.9853	0.7029	0.9255
	33	NTTC-Gin (2914)	0.307	11	180	3.53	0.9209	0.9565	0.7479	0.8561
	21	Sprint (1239)	0.298	12	216	3.07	0.9718	0.9812	0.7470	0.8557
	13	Level-3 (3356)	0.169	25	378	4.49	0.2708	0.9393	-0.0918	0.7982
	12	-/-	0.149	28	436	4.98	0.2055	0.9381	-0.1217	0.7392
	20	Sprint (1239)	0.287	16	528	3.13	0.9866	0.9928	0.5805	0.8488
	9	-/-	0.251	13	741	3.29	0.9901	0.9930	0.7149	0.8622
	Transit	40	JanetUK (786)	0.132	14	336	2.69	0.9714	0.9825	0.8049
45		Iunet (1267)	0.246	11	598	3.88	0.8506	0.9468	0.8887	0.9688
38		-/-	0.231	12	645	3.75	0.8790	0.9516	0.9094	0.9568
39		-/-	0.038	13	711	3.45	0.9470	0.9826	0.5354	0.9536
44		Telecom Italia (3269)	0.037	13	995	3.65	0.7950	0.9828	0.3362	0.8699

topologies. No regularities are reported; their diameter progressively increases while their clustering coefficient decreases, as the network 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), which is worth some more comments. In the subsection III-C we elaborate on the properties of this ISP topology and study the way they affect the considered correlation debate.

2) *CBC vs. egoCBC*: We devote the last set of experiments to the assessment of the ego variant of the CBC metric. The high ρ values of Table V suggest significant positive rank correlation in all considered ISP topologies. Especially for the outlier case of Level-3 networks the correlation turns out to be considerably increased, compared to the one between the BC variants (see Table V). 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. Table V practically *recommends* the application of the egoCBC metric in networking mechanisms.

C. Pathologies in the Level-3 ISP snapshots

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

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 $r_{P_{rs}}$ ones. To further elaborate on this, we studied 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. 4). This finding seems consistent with the remark about the role of nodes with degree one in [14]. Nodes with degree 1 have zero values of both BC and egoBC. Therefore, a network availing a large percentage of such nodes has higher

TABLE VI
CORRELATION STUDY (SPEARMAN COEFFICIENT) BETWEEN CBC-EGOCBC ON INTRA-DOMAIN ISP TOPOLOGIES

DataSet	ISP (AS number)	<Clustering coeff.>	Diameter	Size	<degree>	CBC vs. ego-CBC		
						ego-net. $r=1$	95% Conf. Inter.	
Tier-1	36	Global Crossing (3549)	0.546	10	76	3.71	0.9568	0.008
	35	-/-	0.479	9	100	3.78	0.9489	0.013
	33	NTTC-Gin (2914)	0.307	11	180	3.53	0.9554	0.003
	21	Sprint (1239)	0.298	12	216	3.07	0.9824	0.002
	13	Level-3 (3356)	0.169	25	378	4.49	0.7336	0.007
	12	-/-	0.149	28	436	4.98	0.7035	0.005
	20	Sprint (1239)	0.287	16	528	3.13	0.9847	0.003
	9	-/-	0.251	13	741	3.29	0.9884	0.002
Transit	40	JanetUK (786)	0.132	14	336	2.69	0.9819	0.001
	45	Iunet (1267)	0.246	11	598	3.88	0.7825	0.033
	38	-/-	0.231	12	645	3.75	0.8062	0.022
	39	-/-	0.038	13	711	3.45	0.9370	0.016
	44	Telecom Italia (3269)	0.037	13	995	3.65	0.9902	0.001

chances for an overall higher correlation coefficient than those with fewer nodes. To test further this hypothesis, we generated multiple random graphs [18] feeding the generator with the degree distribution of the Level-3 outliers. Precisely, for each one of the Level-3 16, 18, 27 and 28 datasets 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 VII 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 V) improves significantly when egoBC is computed within the second-order ego network.

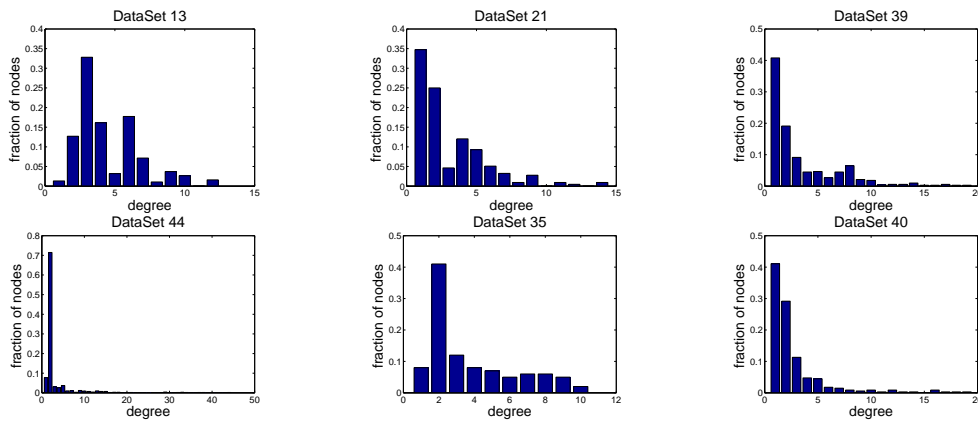


Fig. 4. Indicative set of ISP networks' degree distribution

Following the remark about nodes with degree one we present in Table VII 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

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

DataSet	ISP (AS number)	$\langle CC \rangle$	Diameter	Size	$\langle \text{degree} \rangle$	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	<i>generated</i>	0.0244	8	-/-	4.95	6.34	0.9580	0.93	0
16b	<i>generated</i>	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	<i>generated</i>	0.0516	6	-/-	4.23	3.79	0.8495	1.92	0
18b	<i>generated</i>	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	<i>generated</i>	0.0102	10	-/-	3.94	3.22	0.9053	1.48	0
27b	<i>generated</i>	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	<i>generated</i>	0.0074	10	-/-	4.09	3.77	0.9069	1.14	0
28b	<i>generated</i>	0.0169	10	-/-	4.03	3.49	0.9229	2.00	0

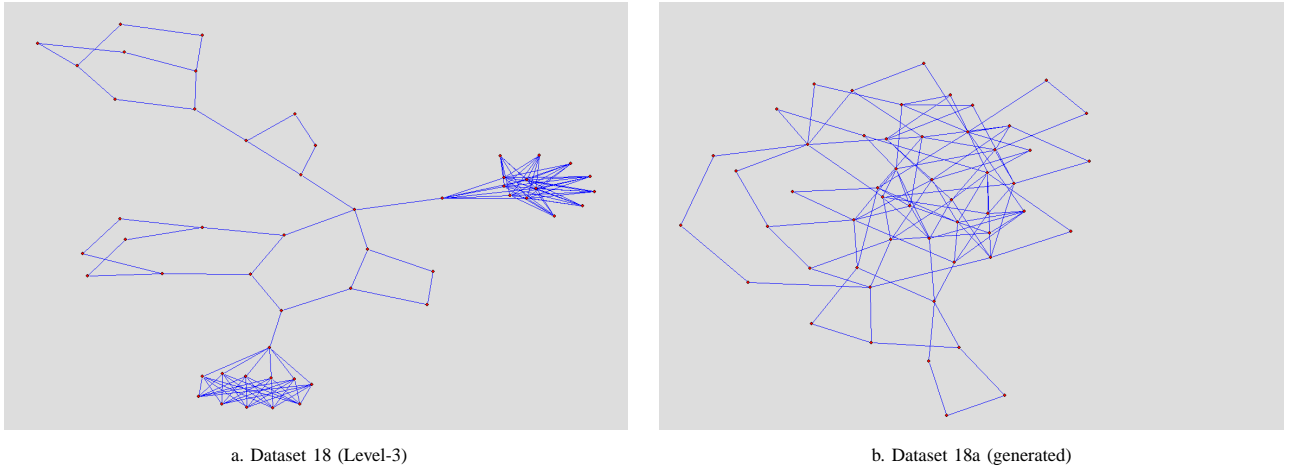


Fig. 5. Visualization of an original Level-3 snapshot (a) and a generated topology with the same degree distribution (b).

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.

To shed some light on the topology configuration that causes this counterintuitive result, we visualize in Figure 5 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. 6a or 5, 6 and 7 in Fig. 6b are the ones that fall in this category and therefore, exhibit higher ego- than socio- BC values. In particular node 5 in Fig. 6a 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

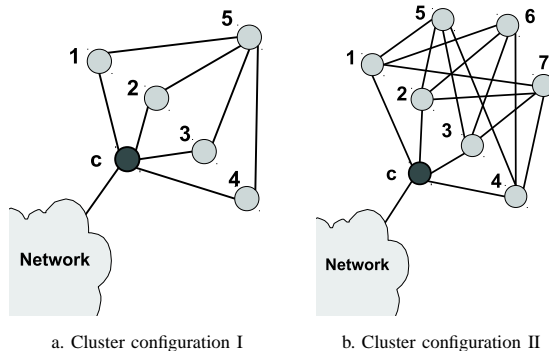


Fig. 6. Configurations of topology’s cluster endings that result in egoBC values of certain nodes exceeding their BC ones.

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.

D. Applications of egocentric metrics

We now outline two examples, where, given their close correlation with sociocentric centrality metrics, their locally measured variants can assist with network operations and facilitate the practical implementation of networking protocols.

Firstly, we sketch the use of a locally computed centrality metric for proactive fault detection. We consider a large-scale wireless sensor network that performs distributed environmental sensing. Such networks consist of several inexpensive nodes, which are scattered in the sensor field and operate under severe energy and computational power constraints [19]. Collected information is routed to a single data sink node. Given the location of the sink s , each sensor node u can easily compute its egoCBC(u ; s) value (see Section II). Local comparison of values or a ranking carried out more centrally after these values are communicated to the sink node would reveal the globally most central nodes with respect to the sink location; hence, those nodes that are about to bear most of the many-to-one information flows towards the sink and accordingly, a significant portion of the overall message load. This piece of information can be directly used to roughly estimate the energy consumption of the sensors or can motivate and calibrate a load-balancing mechanism.

Secondly, we consider the use of localized centrality metrics to facilitate scalable migration of service facilities. The aim is to eventually minimize the service access cost within distributed environments with self-organization characteristics. We have proposed a scalable heuristic that converts the original service placement problem into a few smaller ones [6]; the service migrates towards cost-effective locations by iteratively solving small-scale optimizations on carefully selected subsets of key-nodes. These key nodes are selected on the basis of their centrality, measured

locally via egocentric measurements. The full service placement protocol is outlined in [8].

IV. RELATED WORK

Related work in literature could be organized under two threads. The first one, to which this paper is attached, questions *whether and how much correlation exists* between locally and globally computed centrality metrics; the main contributions hereto come from the area of sociology. Experimental evidence for positive correlation between sociocentric and egocentric BC is given in [20]. The evaluation includes various social networks of sizes ranging from 14-217 nodes and in all cases high *Pearson's correlation* values are reported. The author proposes an explanation for the deviation of different nodes from the respective regression line based on their hub/bridging functionality within the network. Similar positive conclusions, this time for synthetic random graphs $\mathcal{RG}(N, p)$ of size $N \in [25, 500]$ and edge occurrence probability $p \in [0.1, 0.6]$, are drawn in [14]. The mean Pearson correlation coefficient ranges from 0.85 to 0.98, its variance becoming smaller with larger graph sizes. The authors conclude that the correlation is stronger when the network nodes have either very similar or very differentiated sociocentric BC scores.

Closer to our work in terms of topology (large AS-level snapshots) is [10], but reports a linear behavior of the mean BC metric against the node degree. In order to study the scale-free properties of the Internet, the authors analyze the probability distribution of various quantities including betweenness centrality for different time snapshot of Internet maps. They observe that these distributions are characterized by scaling exponents that are stationary in time. Moreover, they argue that the nontrivial betweenness and connectivity correlation is due to the Internet hierarchical structure. Similar power-law behavior has been reported by Goh *et al.* [11] for the *load* distribution in a scale-free network (note that the *load* of node k is only a betweenness centrality equivalent; it is the total amount of data packets passing through k when all pairs of vertices send and receive a data packet unit along the shortest paths connecting the pair). 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 study from networking community [21], compares the ego- and sociocentric counterparts of the recently proposed bridging centrality, which identifies the centrality of a node between highly connected regions. They experiment with one synthetic, one social and one wireless mesh network of sizes 11, 14 and 8 nodes, respectively, and find that both metrics rank nodes similarly. To the best of our knowledge, our study is the first that looks into both synthetic and real world network topologies of sizes up to 1000 nodes. Moreover, it expands the definition of egocentric metrics to gain a more thorough view of the underlying complexity *vs.* accuracy tradeoff.

The second thread assumes a positive reply to the first question and explores *how could this correlation be exploited* in the practical specification of networking protocols. The study in [22] argues in favor of the utility of localized centrality metrics in identifying users with rich social network within large-scale collaborative networks. The author experiments with a commercial collaborative bookmarking solution studying the graph of implicit (shared

URL tags) and explicit (subscriptions) relationships among users. Insights from social network analysis have been more directly applied in the area of Delay Tolerant Networks. Both [4] and [5] draw on egocentric estimates of nodes' betweenness centralities to derive new forwarding protocols that, when correctly tuned, can significantly improve performance over more naive approaches.

V. CONCLUSIONS

We have placed node betweenness centrality metrics against their egocentrically computed variants on a wide range of both synthetic and real-world network topologies. Egocentric computations come at considerably lower cost than their sociocentric counterparts. Therefore, the reported strong correlation between the two variants, in particular for real world ISP topologies, lays an argument in favor of the feasibility of socioaware networking.

REFERENCES

- [1] J. Kleinberg, "The convergence of social and technological networks," *Communications of the ACM*, vol. 51(11), pp. 66–72, 2008.
- [2] B. Wellman, "Computer networks as social networks," *Science*, vol. 293, pp. 2031–2034, September 2001.
- [3] S. Wasserman and K. Faust, *Social network analysis: Methods and applications*. Cambridge Univ Pr, 1994.
- [4] 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.
- [5] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: social-based forwarding in delay tolerant networks," in *ACM MobiHoc '08*, China, 2008.
- [6] P. Pantazopoulos, M. Karaliopoulos, and I. Stavrakakis, "Scalable distributed service migration via complex networks analysis," Tech. Rep. [Online]. Available: <http://cgi.di.uoa.gr/~istavrak/publications.html>
- [7] L. C. Freeman, "Centered graphs and the structure of ego networks," *Mathematical Social Sciences*, no. 3, pp. 291–234, 1982.
- [8] P. Pantazopoulos *et al.*, "Efficient social-aware content placement for opportunistic networks," in *IFIP/IEEE WONS*, Slovenia, 2010.
- [9] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.
- [10] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, "Large-scale topological and dynamical properties of the internet," *Phys. Rev. E*, vol. 65, no. 6, p. 066130, Jun 2002.
- [11] K.-I. Goh, B. Kahng, and D. Kim, "Universal behavior of load distribution in scale-free networks," *Phys. Rev. Lett.*, vol. 87, no. 27, Dec 2001.
- [12] S. P. Borgatti, "Centrality and network flow," *Social Networks*, no. 27, pp. 55–71, 2005.
- [13] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, vol. 25, pp. 163–177, 2001.
- [14] M. Everett and S. P. Borgatti, "Ego network betweenness," *Social Networks*, vol. 27, no. 1, pp. 31–38, 2005.
- [15] J.-J. Pansiot, "mrinfo dataset." [Online]. Available: <http://svnet.u-strasbg.fr/mrinfo/>
- [16] J.-J. Pansiot *et al.*, "Extracting intra-domain topology from mrinfo probing," in *Proc. PAM*, Zurich, Switzerland, April 2010.
- [17] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.
- [18] T. Britton *et al.*, "Generating simple random graphs with prescribed degree distribution," *J. STAT. PHYS*, vol. 124, no. 6, 2005.
- [19] I. F. Akyildiz *et al.*, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, pp. 393–422, 2002.
- [20] P. Marsden, "Egocentric and sociocentric measures of network centrality," *Social Networks*, vol. 24, no. 4, pp. 407–422, October 2002.
- [21] S. Nanda and D. Kotz, "Localized bridging centrality for distributed network analysis," in *IEEE ICCCN '08*, Virgin Islands, August 2008.
- [22] E. M. Daly, "Investigation of localised centrality metrics for collaborative networks: What can they reveal?" in *The 20th Irish Conference on Artificial Intelligence and Cognitive Science*, Dublin, Ireland, 2009.

APPENDIX

A. Ego-betweenness computation in grid topologies

Consider a $M \times N$ rectangular grid and let $i, j \in \mathbb{N}$ be the coordinates of a grid node u and d_{ij} its degree. The $egoBC(u, 1)$ value depends solely on the degree of node u . Nevertheless, it is the combination of the grid size and

the position of node u that determines the different possible $egoBC(u, 1)$ values. When we consider the first layer

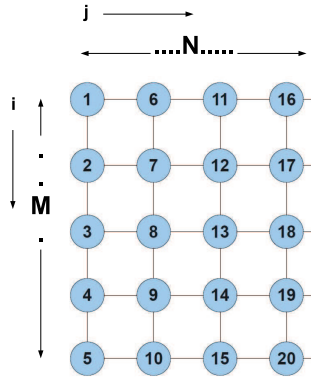


Fig. 7. $M \times N$ rectangular grid

of node u neighbors ($r = 1$) under the constrain of $M, N > 3$, the $egoBC(u, 1)$ can attain three different values given by the following expression:

$$egoBC(u; 1) = \binom{d_{ij}}{2} = \begin{cases} 1, & \text{if } i \cdot j \in \{1, M, N, M \cdot N\} \\ 3, & \text{if } (i-1)(i-M)/(j-1)(j-N) = 0 \text{ or } (j-1)(j-M)/(i-1)(i-N) = 0 \\ 6, & \text{elsewhere} \end{cases} \quad (8)$$