Internet Content Management using Complex Network Analysis techniques

Panagiotis Pantazopoulos*

Department of Informatics and Telecommunications,
National & Kapodistrian University of Athens
Ilissia, 157 84 Athens, Greece
ppantaz@di.uoa.gr

Abstract. With the explosion of (user-generated) content and the heterogeneity of users/devices, today's Internet has evolved into a system of extreme complexity hindering the design of effective network protocols. The availability of global information, either topological or related to the users' profiles is non-realistic if not impossible to obtain. As such, a significant open challenge that the thesis seeks to address involves the management of Internet content in a distributed manner with the utilization of local-scope information. Towards this end, the socio-technical insights provided by the interdisciplinary framework of Complex Network Analysis (CNA) have attracted the interest of the research community, yet remain under-explored. The thesis addresses instances of content management problems over mobile opportunistic networks as well as wired ISP networks at the router-level. Among the mobile opportunistic nodes, we seek to identify the appropriate nodes to store and efficiently provide content to the rest of the network. Over the ISP topologies, we study the distributed Internet service placement and the content search driven by local information. The vulnerability of these networks to node attacks is also explored from both the connectivity and content-related standpoint. The novelty dimension of the thesis lies in the effort to introduce and promote centrality metrics to important parameters for the design of Internet content operations. Our results obtained by analysis and simulations over real-world data, reveal useful insights and provide guidelines as to how (local) centrality information can efficiently drive the management of Internet content.

1 Introduction

Content management is a summary term for a broad range of operations that become relevant from the moment some content is produced somewhere in the communication network infrastructure till the time this content is consumed by end-users, including content search, and discovery, content placement and dissemination. With the role of content continuously growing and the latest advances in Web 2.0 technologies fueling user-generated content, end-users increasingly participate in these operations, departing from the monolithic role of content consumer.

The new role of end-users occasionally acting as content providers as well as the interactions between them and the network devices, traditionally, has not been systematically taken into account in the design of network protocols. The emerging structures

* Dissertation Advisor: Ioannis Stavrakakis, Professor

and properties that are due to these interactions present network environments with "social dimensions" that have recently gained momentum in the context of more user-oriented approaches to data networking. Indeed, there has been recent evidence, which motivates the proposed research, that taking explicit account of these social attributes of end-users does benefit content management operations substantially [2] [3]. In light of the content steadily gaining importance in communication networks, the need to make its management more efficient motivates the introduction of novel, socio-aware approaches. The primary concern is to devise distributed and scalable Internet solutions that utilize locally available information.

The highly interdisciplinary framework of Complex Network Analysis (CNA) seeks to describe the research for modeling the behavior of networked interacting elements. It typically utilizes a set of graph-theoretic metrics for describing relations between nodes or groups of nodes. Expectations within the networking community are that the relevant socio-technical insights could assist simplifying the high networking complexity and benefit the design of efficient protocols. The thesis aims to pursue in-depth this research thread by analyzing pervasive CNA-driven content management mechanisms over the current/future Internet. Content-related operations are nowadays carried out over wired ISP networks as well as over dynamically changing (often time-evolving) topologies such as mobile ad-hoc and opportunistic networks. Those networking settings tend to become highly user-centric network paradigms, realized by the end-user participation. Devising topology- and resource-control mechanisms that accommodate the content management operations is a challenging yet promising task. In particular, the proposed research seeks to exploit the CNA-based knowledge and relevant insights for the optimization of content-centric operations in the context of current and future communication network paradigms. At the same time, studying CNA-concepts over complex real-world Internet topologies, the thesis makes solid contributions to the emerging and much promising scientific discipline of Network Science.

2 State-of-the art and outline of the thesis

In this section we briefly present the related publications and outline the corresponding contents of the thesis.

2.1 Background in opportunistic content placement and data forwarding

The opportunistic communications paradigm Opportunistic networks [15] are self-organizing wireless mobile networks formed by user devices such as PDAs or Smartphones, without requiring any pre-existing network infrastructure. Communications in this environment have been typically supported through Mobile Ad hoc Networks (MANETs). However, MANET solutions work only if a rather stable topology can be established among the nodes, which is often not the case in the presence of users' mobility and low node density. Opportunistic networks support communication among nodes even when no stable multi-hop paths between communication endpoints can be established. In an opportunistic network a node carrying a content addressed to a given destination evaluates if any other node it comes in direct contact is "better suited" than itself to bring the content to the destination. In other words, each contact is opportunistically exploited to bring the content closer and closer to the destination.

Content placement in opportunistic networks Content placement in opportunistic networks can be seen as an instance of the broad data dissemination problem. The latter amounts to the storage of files or, more generally, information objects, in specific network nodes, so that they can be provided to requesting nodes at smaller access costs. Those problems have recently attracted the attention of the networking community. ContentPlace, which is a social-oriented framework for data dissemination constitutes a state-of-the-art relevant solution [1]. ContentPlace assumes that nodes can be aware of the social communities they belong to. Using a general utility-based optimization framework, ContentPlace defines distributed algorithms for nodes to select which content to locally replicate, out of what is available on encountered nodes. These algorithms consider the estimated distribution of content in the network, and the interests of the users with respect to content. Similar approaches have shown that mobility and cooperative content replication strategies can help bridge social groups [11].

In the second Section of the thesis briefly discussed in 3.1, we address the content placement problem in an opportunistic network by incorporating a social dimension to its solution. The proposed algorithm can provide for easy adaptation to dynamic environments due to its local-information-requiring approach and can therefore, cope with complex multimedia content increasingly generated by the network users.

Opportunistic Data Forwarding The opportunistic forwarding problem amounts to the decision that a node has to take on whether some encountered node is more appropriate (or not) to physically carry the data to a given destination. Initial approaches to the problem have essentially been variants of controlled flooding across the network. These schemes reduce the cost of pure epidemic dissemination by setting upper bounds on the message replication(*e.g.*, [18]). More informed decisions are made by forwarding schemes that try to assess the relaying significance (*utility*) of encountered nodes. They may account for the frequency of encounters with the destination node or more general social context such as preferably visited places, common to the candidate relay and the destination node. Recently, social information has been introduced into the node relaying utility functions through direct reuse of Complex Network Analysis concepts. Examples of this approach are the SimBetTS and BubbleRap protocols. In both cases, the CNA metrics are computed over contact graphs that effectively aggregate the sequence of node encounters over certain time windows. In SimBetTS [3] the nodes' utilities are sums of their centrality, similarity, and tie strength values, the latter reflecting the frequency, duration, and recency of the contacts with other nodes. Whereas, BubbleRap [4] explicitly assumes the existence of nodes' communities and manipulates the centrality of encountered nodes, both within their communities and globally across the whole network, to route messages within and across these communities, respectively. Both protocols have reported enhanced performance over the controlled flooding and identified centrality as the metric with the dominant impact on routing even when it is combined with other social metrics.

What we seek to assess in the third Section of the thesis, summarized here in 3.2, is the inherent weaknesses of centrality-based routing that pose hard limits to the performance of socioaware opportunistic routing. Relevant questions amount to but are not limited to: How close-to-optimal can routing decisions based on centrality metrics be? How do different alternatives for *computing* node centrality affect routing performance?

2.2 Background in content/service placement over physical topologies

Efficient content/service placement can dramatically reduce access speeds and related costs, whether viewed in technical or economical terms, and, therefore, improve the quality of the provided service. The optimal placement of content (or a service facility) within a network structure has been typically tackled as an instance of the facility location problem [7]. Input to the problem is the topology of the network nodes that may store the content or (host the service), their costs of installation and the distribution of service demand across the network users. An appropriate formulation of the problem is needed to cope with the current networking trends.

Nowadays, the Web2.0 technologies have enabled a paradigm shift towards more user-centric approaches to content generation and provision. This shift is strongly evidenced in the abundance of *User-Generated Content*(UGC) in social networking sites or video distribution sites (*e.g.*, YouTube). The generalization of the UGC concept towards services has become a major trend in user-oriented networking. The user-oriented service creation concept engages end-users in the generation and distribution of service components. In parallel with the proliferation of the so-called *User-Generated Service* (UGS) paradigm, significant research is being carried out on the design and deployment of energy-efficient data storage architectures. Numerous peer devices such *e.g.*, home-gateways are instrumented using virtualization to create a distributed Internet service platform that leverages end-user proximity. The "in-network storage" argument has also been a key-concept of the emerging Information-centric networking paradigm [2]. In the intersection of these trends, we anticipate a rich ecosystem of service instances that will be generated in almost every network location and will have access to storage resources in various network locations. The technical challenge then is how to place them in a way that minimizes their access cost.

To identify the optimal location a large optimization problem needs to be solved. Thus, approximation algorithms have been so far proposed, the majority of which address the problem employing global knowledge. The Greedy algorithm [16] is a relevant heuristic solution that sequentially places one replica at a time; the current one is placed at the lowest-cost location exhaustively determined under the assumptions that a) the so far placed replicas remain fixed b) a node's requests are directed to the closest replica. The approach achieves placements within a factor of 1.1-1.5 of the optimal for synthetic and real-world network topologies under real-world demand patterns. However, in distributed settings it is infeasible to acquire global input information. The problem is further amplified by the massive *UGC*/*UGS* trend that especially calls for scalable solutions. A few distributed approaches have been so far proposed; most of them are heuristics that employ locally-determined information and migrate the service towards prominent locations. The authors in [17] propose the R-ball heuristic. They iteratively solve multiple small-scale optimization problems within an area of R-hops from the current location of each service facility and move each service towards (near-)optimal locations. A slightly different instance exploits the shortest-path tree structures induced on the network graph by the routing protocol operation to estimate upper bounds for the aggregate cost when the service migrates to its immediate neighbors.

In the forth Section of the thesis we present a scalable heuristic approach to deal with the complexity and limitations of the distributed service placement. Node centrality insights help us iteratively migrate service facilities towards near-optimal locations achieving very good accuracy and fast convergence. A brief description of the algorithm and the corresponding protocol implementations can be found in this document, in 4.1.

2.3 Background in local centrality metrics

The analytical (graph-theoretic) tools that the Complex Network Analysis (CNA) provides are expected to benefit the design of efficient network protocols. Indeed, there has been evidence that CNA insights can improve network functions such as content-caching strategies in wired networks [2] and routing/forwarding in opportunistic networks [3]. Common denominator to these efforts is the use of CNA-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, if not infeasible, in large-scale networks such as the Internet ISP topologies. 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 metrics have, in fact, found their way into protocol implementations [3], [2]. Nevertheless, the capacity of these local approximations to substitute the globally computed sociocentric metrics over the Internet is almost always taken for granted rather than evaluated.

In Section 5 of the thesis, summarized in 4.2, we employ ISP network topologies and question how well do sociocentric node centrality metrics, computed under global topological information, correlate with their egocentric variants, as computed locally over the nodes' ego networks. More importantly, 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 some elementary network operations *i.e.*, instances of content-related Internet protocols.

3 Content Management over Opportunistic Networks

In this section we briefly describe the solutions proposed in the thesis regarding the placement of content as well as the message forwarding over opportunistic networks.

3.1 Content Placement

CNA insights are being exploited here to determine the optimal physical location of the node to host some content by solving iteratively a spatially restricted, low-cost

1-median problem (instead of a costly, global one) [11]. Consequently, the proposed algorithm can provide for easy adaptation to dynamic environments due to its local-information-requiring approach. As approximations are involved, the achieved solution may not always lead to the identification of the optimal location; nevertheless, the results presented in the thesis show that the divergence error is not substantial.

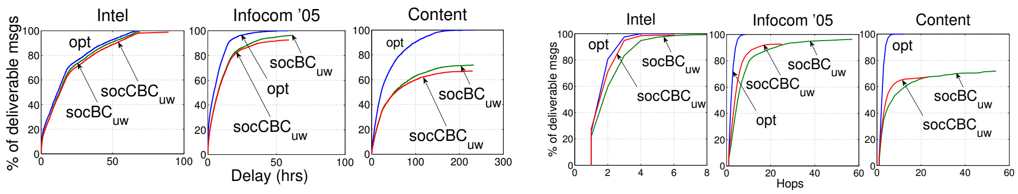
First, the thesis argues that contact patterns among users naturally hint to a graph representation of the network, where a link exists among nodes if they are "enough frequently" in touch with each other. With this representation and the assumption of full topology knowledge over a limited region around some node currently hosting the content, the thesis introduces a CNA-based criterion, for selecting a number of fairly neighboring nodes (forming a subgraph) to take part in a small-scale, local solution. This, may be derived by employing any well-studied centralized approach. The selected nodes (a percentage of total number) are the ones, having the top values of an innovative metric, inspired by typical CNA centrality metrics, that plays a twofold key role. First, it captures a node's significance, regarding its capability to transport content efficiently. Second, it captures the contribution (to the *1-median* solution) of incoming demand from the rest of the network nodes, not included in the above subgraph. In mathematical terms the introduced metric called Conditional Betweenness Centrality (CBC), captures the topological centrality of a network node with respect to a specific node t. If σ_{st} denotes the number of shortest paths between any two nodes s and t in a connected graph G=(V, E) and $\sigma_{st}(n)$ is the number of shortest paths passing through node $n \in$$V$, then CBC is:

$$CBC(u;t) = \sum_{s \in V, u \neq t} \frac{\sigma_{st}(u)}{\sigma_{st}} \qquad (1)$$

The thesis shows that the demand, following a uniform model, can be deduced from this metric which is a measure of the role of every node in the information flow, towards the one having the content. After solving the small-scale $1 - median$ problem, the node that minimizes the content provision cost (if hosting the content) is identified. Then, the content is assumed to be placed in this node. The new subgraph is determined around this node and the new small-scale optimization problem is solved. This iterative procedure repeats until no further movement of the content is possible. The content therefore, moves according to the outcome of the optimization, on a cost-decreasing path, trying to reach the optimal location. In the thesis we analyze the performance of the proposed heuristic studying its degree of approximation of the optimal centralized solution on two types of graphs representing different user contact patterns, i.e. Erdős-Rényi and the Barabási-Albert graph model. Our simulation shows that the heuristic achieves satisfying convergence to the optimal solution. Interestingly, it performs much better under the B-A model, which is known to capture user social interactions and contact patterns.

3.2 Message Forwarding

In the thesis we present our study of centrality-based opportunistic routing employing real traces of pairwise node encounters. We have used five well-known experimental traces, gathered in the context of the Haggle Project [8]. Rather than consid-



a. b.

Fig. 1. a) Message delivery delay distribution for destination-aware ($socCBC_{uw}$) and -unaware ($socBC_{uw}$) centrality-based forwarding. b) Message hopcount distribution for destination-aware ($socCBC_{uw}$) and -unaware ($socBC_{uw}$) centrality-based forwarding.

ering a particular protocol, we have assessed the basic routing primitive of centrality-based schemes. We have studied three fundamental alternatives to the sociocentric (*i.e.*, global-info-requiring) computation of the betweenness centrality over contact graphs. Methodologically, our approach amounts to the following steps: First we transform the encounters' history to a graph, over which the node centrality values are computed. We produce two different static graph representations for each trace: one unweighted and one weighted, where the edge weight equals the inverse of their encounters' count. We then generate message triplets $msg(s, d, t)$, where the message source s, destination d, and generation time t are randomly chosen, and emulate their paths over the traces [5]. As the trace is replayed (sequentially read), network nodes compute online their BC values and make forwarding decisions for each message. The generated $msg(s, d, t)$ messages are routed to their destination through successive 'greedy' forwarding decisions. More specifically, if the message is with node u, then u will forward the message to another node, say k, upon its next encounter with it, as long as its betweenness centrality value $BC(k)$ is higher than its own BC value, $BC(u)$ We consider five possible ways to compute BC values assuming that nodes avail perfect information about the history of encounters in the network. Hence, any routing performance penalty is due to the (lack of) informative power of the metric rather than information unavailability. The various BC computations are marked by the abbreviation terms $soc(C)BC_{uw}$, $egoBC_{uw}$, $socBC_w$, $egoBC_w$, for routing based on global (*i.e.*, sociocentric) and local (*i.e.*, egocentric) (C)BC values estimated on unweighted or weighted graphs, respectively. We measure the message delivery delay and the number of message forwarding hops.

A representative result in Fig. 1 compares the performance of optimal, BC- and CBC-based routing schemes over unweighted contact graphs. As expected, the centrality-based approaches perform worse than the optimal method both in terms of message delay and hops. Up to 30% of messages are trapped and do not reach their destination. Note that the performance lag of centrality-based schemes varies from trace-to-trace and heavily depends on the extent that the mobility patterns of users mix with each other. Less intuitively, the CBC-based forwarding does not outperform the BC-based forwarding with respect to delays. Although the use of CBC implies message routing through more appropriate relays towards the destination, there are numerous graph instances where its interpretation turns out to be problematic. When the aggregation of contacts yields non-connected clusters of nodes, the CBC values of nodes outside the destination's cluster are by default zero. Messages stay long with the source node or are trapped quickly at some intermediate node. On the contrary, with use of the BC metric, nodes take on easier non-zero values and the resulting variance of the metric across nodes lets the message hop from one node to the other. On the other hand, this extra message agility comes at higher cost; under BC-based forwarding, messages end up traversing up to 50% longer paths than under the use of CBC (Fig 1.b), *i.e.*, messages end up traveling far more in the network.

4 Content management over ISP network topologies

In this section we describe the solutions proposed in the thesis regarding the distributed service placement and the utility of local centrality metrics when used to drive basic network operations.

4.1 The Distributed Service Placement Problem

Given the network topology and demand dynamics, the k-$median$ problem prescribes the locations for instantiating a fixed number of service facilities so as to minimize the aggregate cost of accessing them over all network users [7]. We focus on the single facility scenario that matches better the UGS paradigm, *i.e.*, various service generated in the network raising small-scale interest so that replication of their facilities be less attractive [14]. The network is represented by an undirected connected graph $G(V, E)$ of $|V|$ nodes and $|E|$ edges. If $w(n)$ denotes the aggregate service demand generated by node n and $d(k, n)$ is the minimum cost path between nodes k and n, then the 1-median problem formulation seeks to minimize the access cost of a service located at node $k \in V$: $Cost(k) = \sum_{n \in \mathcal{V}} w(n) \cdot d(k, n)$.

The proposed heuristic (called cDSMA) relies on a traffic-aware variant [12] of the earlier introduced CBC metric (eq.1) that we call *weighted*-CBC ($wCBC$). It helps each service instance to migrate towards its final location through a finite number of steps:

Step 1*: Initialization*. The algorithm execution starts at the node s that initially generates the service facility. The service placement cost at node s is assigned an infinite value to secure the first iteration

Step 2*: Metric computation and 1-median subgraph derivation*. Next, the computation of $wCBC(u; s)$ metric takes place for every node u in the network graph $G(V, E)$. Nodes featuring the top $\alpha\%$ $wCBC$ values together with the current service $Host$ form the subgraph G^i_{Host} (i^{th} iteration), over which the 1-median problem will be solved.

Step 3*: Mapping the demand of the remaining nodes on the subgraph*. By restricting the solution domain to the G^i_{Host} subgraph, the contribution of the "outside world" to the service provisioning cost would be totally neglected. To allow for its inclusion, the demand for service from the $G \setminus G^i_{Host}$ nodes is mapped on the G^i_{Host} ones.

Step 4*: 1-median problem solution and service migration to the new host node*. Any centralized technique may be used to solve this small-scale optimization and determine

Table 1. cDSMA accuracy assessment for various datasets under different demand distributions

ISP	AS number	Nodes	Diameter	Mean Degree	Minimum $\|G^i_{Host}\|$ nodes to yield cost$\leq 1.025 opt$ Uniform demand	Zipf demand (s=1)
Global Crossing	3549	100	9	3.78	7	6
NTTC-Gin	2914	180	11	3.53	18	10
Sprint	1239	184	13	3.06	8	7
DFN-IPX-Win	680	253	14	2.62	7	6
Level-3	3356	378	25	4.49	4	4

the optimal location of the new $Host$ among the G^i_{Host} nodes. If the value of the emerging cost is smaller than the current one, the service is moved to the new $Host$ and the algorithm iterates steering the service to (near-)optimal locations.

Practical protocol implementations for different routing strategies The $wCBC(u;t)$ metric practically represents the service demand that node u routes towards node t, including its own demand and the transit demand flowing from other network nodes through u towards t. Therefore, individual nodes may, in principle, *estimate* their own metric values $wCBC$ through passive measurements of the service demand they route towards the current service host node. What is actually computed theoretically demanding global information about the network topology and service demand, can be locally approximated by u providing the basis for the practical implementation of a distributed solution. The approximation lies in the fact that what is measured, even with perfect accuracy, is not always equal to the nominal $wCBC$ value. In the thesis we show that what matters is the estimate of the actual demand routed through the node.

Using this locally-obtained estimate values of each node's $wCBC$ the thesis proposes a real-world *distributed* implementation for cDSMA, catering for all challenges related to distributed operation: how the node each time hosting the service collects topological and demand information and how it uses it to reconstruct the inputs needed by the algorithm. Each network node communicates the $wCBC$ estimates via dedicated messages to the current service host. This information can then be processed by the service host to extract partial topological information about the 1-median subgraph and determine the next service host on its migration path.

Results on ISP topologies We present a portion of our experimental results regarding the performance of the theoretical solution *i.e.*, when the involved metrics are globally determined. We employ real-world ISP network topologies [10] that feature adequate variance in size, diameter, and connectivity degree statistics. Table 1 reports the minimum number of nodes $|G^i_{Host}|$ required to achieve a solution that lies within 2.5% of the optimal for different levels of service demand asymmetry. The $|G^i_{Host}|$ values show remarkable insensitivity to both topological structure and service demand dynamics. Employing 4.5% of the total number of nodes or 6% for the least favorable case suffices to obtain very good accuracy across all ISP topologies. Likewise, the required 1-median subgraph size remains in almost all experiments practically invariable with the demand distribution skewness.

More importantly, the thesis shows that the cDSMA heuristic lends to realistic protocol implementations that realize the distributed service placement. Our relevant results suggest that the proposed practical implementations preserve similar advantages with the theoretical cDSMA algorithm regardless of the employed routing policies. Our experiments also show that the cDSMA implementations exhibit welcome scalability properties with respect to the number of the involved nodes and message overhead. Finally, the concentration of Internet services (over highly central nodes) is almost negligible when the cDSMA implementations operate under realistic service demand scenarios.

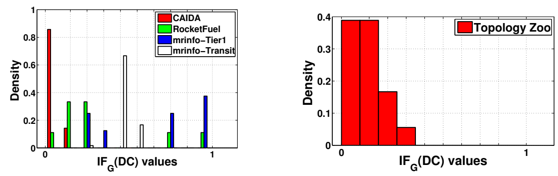
4.2 The utility of local centrality metrics

A realistic alternative for assessing node centrality yet suitable for distributed Internet environments draws on local structures such a node's ego network [6]; the latter consists of the subgraph involving the ego-node, its 1-hop neighbors, and their inter-connections. Egocentric measurements, carried out within their immediate locality, let nodes derive local approximations of their centrality.

In view of the fact that most centrality-driven protocol implementations rely on the (BC-determined) ranking of nodes rather than the absolute values we first investigated how well do BC metrics computed under global topological information, correlate with their locally-determined variants over real-world Internet topologies [13]. Our results presented in Chapter 5 of the thesis suggest that the original BC and its destination-aware variant, exhibit high rank-correlation (in the order of 0.8-0.9) with their local counterparts across almost all 20 ISP snapshots studied. 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 have 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 content search scheme employing local centrality metrics produce significantly different navigation patterns and lower hitrates, 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.

5 Vulnerability of ISP networks to intelligent node attacks

In the sixth Section of the thesis we have studied a problem that relates implicitly to the Internet content management; namely, we have experimentally assessed how vulnerable are the router-level Internet topologies to attacks (*i.e.*, removals) targeting their top central nodes [9]. Content-related operations will benefit if these topologies can maintain high levels of network connectivity and volumes of accommodated traffic, in face of the attacks. To identify the top central nodes of each topology we have used the seven most popular centrality metrics and first studied how different are the node rankings (measured by correlation and top-k overlap), they induce. Included in the set of these centralities, is the only local metric *i.e.*, the degree centrality(DC) which is of apparent interest since it requires almost no computation to be obtained. Then, with respect to those rankings, highly central nodes are removed and the impact on both the connectivity and the traffic-carrying capacity of each network are assessed.



a. b.

Fig. 2. Empirical probability mass function of the normalized distance measure $IF_G(DC)$ computed *w.r.t* the size of the giant component (a) and the maximum accommodated flow (b) for different ISP network datasets.

In terms of the matching of node rankings, we have found (inline with intuition) that it is the top-5% overlap rather than the full rank correlation that predicts more accurately when the node removals that are determined by two different indices have similar impact on the network. Regarding the centrality-driven attacks, the use of different centralities for the choice of the nodes to-be-attacked varies significantly their impact on the network. Focusing on the only local metric (*i.e.*, DC) we have introduced a normalized distance metric called IF to quantify how close to the most catastrophic attack is the impact of the DC-driven removals. Our results suggest that the locally computed DC approximates closely the most catastrophic global centrality metrics with respect to the traffic-carrying capacity (Fig. 2.b). On the contrary, such an approximation is more topology-dependent in terms of connectivity (Fig. 2.a).

6 Conclusions

The thesis proposes algorithms and protocols to enhance conventional content (or service) management functions over the Internet. The task of content management may include the generation, transfer and provision of some piece of various-typed informative content over underlying networks of diverse characteristics. The relevant research is triggered by both the high complexity of emerging distributed Internet environments as well as the preliminary yet promising results of the upcoming Complex Network Analysis (CNA) framework. Employing centrality metrics, the thesis seeks to devise content management solutions that can cope with today's content explosion as well as the need for distributed approaches utilizing local information. The focus lies on two instances of the current/emerging Internet environment, the mobile opportunistic networks and the wired ISP router-level topologies. In the former case, we seek to a) identify the nodes to store a content item that will be provided to the rest of the network nodes at minimum access cost and b) experimentally assess the impact of centrality computations on the efficacy of socio-aware opportunistic forwarding. In the latter case, the question investigated pertains to the distributed placement of Internet services and the utility of local centrality metrics to drive basic network operations. Finally, we study CNA metrics in the Internet vulnerability context. The thesis introduces and promotes centrality metrics to important parameters for the design of Internet content-related operations. Our results obtained by analysis and simulations over real-world data, essentially provide yet another evidence that the incorporation of complex network structures information is a promising path to the development of efficient content-centric networking protocols for future Internet users.

References

1. Boldrini, C., Conti, M., Passarella, A.: Design and performance evaluation of contentplace, a social-aware data dissemination system for opportunistic networks. Computer Networks 54(4), 589–604 (2010)
2. Chai, W.K., He, D., Psaras, I., Pavlou, G.: Cache "less for more" in information-centric networks. In: Proc. of the 11th IFIP Networking. Prague, Czech Republic (May 2012)
3. 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)
4. Hui, P., Crowcroft, J., Yoneki, E.: Bubble rap: Social-based forwarding in delay-tolerant networks. IEEE Trans. Mob. Comput. 10(11), 1576 –1589 (nov 2011)
5. Karaliopoulos, M., Pantazopoulos, P., Jaho, E., Stavrakaki, I.: Trace-driven Analysis of Data Forwarding in Opportunistic Networks. In: Proc. of the 2nd Conference on the Analysis of Mobile Phone Datasets and Networks (NetMob'11). Cambridge, MA, USA (2011)
6. Marsden, P.: Egocentric and sociocentric measures of network centrality. Social Networks 24(4), 407–422 (October 2002)
7. Mirchandani, P., R.Francis: Discrete location theory. John Wiley and Sons (1990)
8. Nikolopoulos, P., Papadimitriou, T., Pantazopoulos, P., Karaliopoulos, M., Stavrakakis, I.: How much off-center are centrality metrics for routing in opportunistic networks. In: Proc. of the 6th ACM Workshop on Challenged Networks. CHANTS '11, Las Vegas, USA (2011)
9. Nomikos, G., Pantazopoulos, P., Karaliopoulos, M., Stavrakakis, I.: Comparative assessment of centrality indices and implications on the vulnerability of ISP networks. In: 26th International Teletraffic Congress (ITC 2014). Karlskrona, Sweden (Sep 2014)
10. Pansiot, J.J., et al.: Extracting intra-domain topology from mrinfo probing. In: Proc. PAM. Zurich, Switzerland (April 2010)
11. Pantazopoulos, P., Stavrakakis, I., Passarella, A., Conti, M.: Efficient social-aware content placement for opportunistic networks. In: IEEE WONS. Kranjska Gora, Slovenia (2010)
12. Pantazopoulos, P., Karaliopoulos, M., Stavrakakis, I.: Centrality-driven scalable service migration. In: The 23rd International Teletraffic Congress (ITC). San Francisco, USA (2011)
13. Pantazopoulos, P., Karaliopoulos, M., Stavrakakis, I.: On the local approximations of node centrality in internet router-level topologies. In: IFIP IWSOS. Mallorca, Spain (Mar 2013)
14. Pantazopoulos, P., Karaliopoulos, M., Stavrakakis, I.: Distributed placement of autonomic internet services. IEEE Trans. on Parallel and Distributed Systems 25(7), 1702–1712 (2014)
15. Pelusi, L., Passarella, A., Conti, M.: Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. IEEE Communications Magazine 44(11), 134 (2006)
16. Qiu, L., Padmanabhan, V.N., Voelker, G.M.: On the placement of web server replicas. In: Proceedings of the IEEE INFOCOM'01. vol. 3, pp. 1587–1596 (2001)
17. Smaragdakis, G., et al.: Distributed Server Migration for Scalable Internet Service Deployment. IEEE/ACM Transactions on Networking 22(3) (2014)
18. Spyropoulos, T., et al.: Efficient routing in intermittently connected mobile networks: The Multiple-Copy case. IEEE/ACM Trans. Netw. 16(1), 77–90 (Feb 2008)