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

Panagiotis Pantazopoulos Merkourios Karaliopoulos
Ioannis Stavrakakis

Department of Informatics and Telecommunications
National & Kapodistrian University of Athens

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

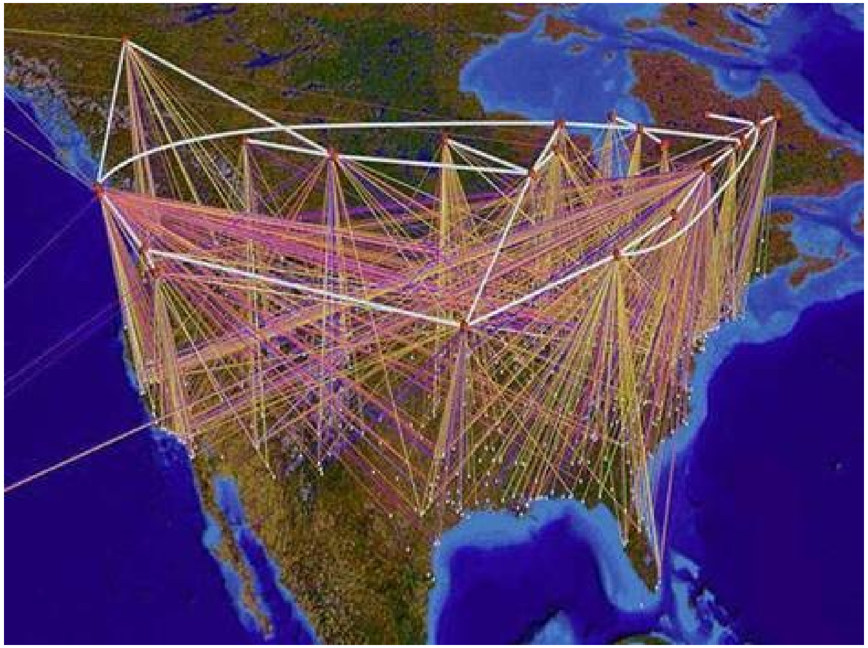
IWSOS, 9-10th of May, 2013
Palma de Mallorca

Social Network Analysis Toolbox in Networking

Social Network Analysis Toolbox in Networking

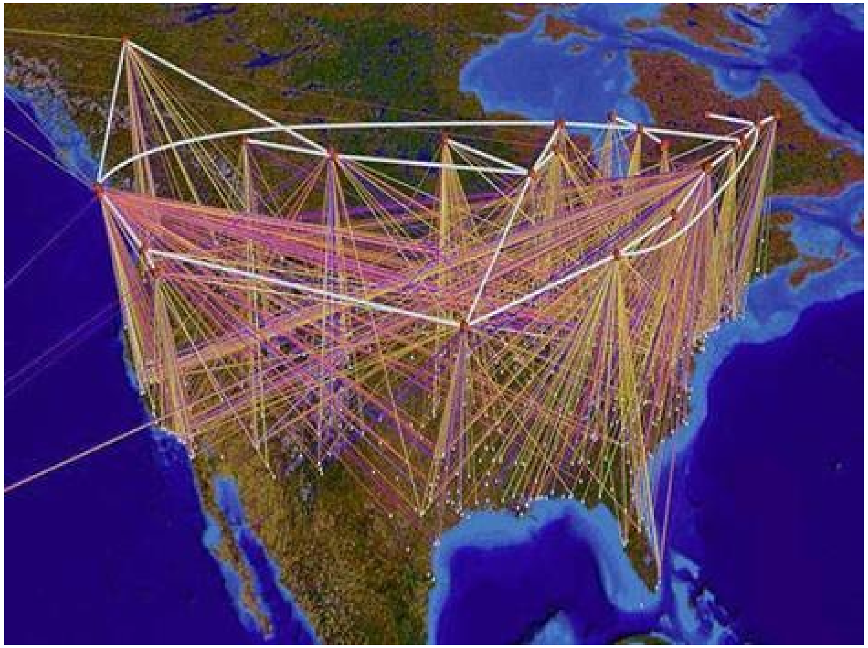
- Computer networks : systems of increasing complexity

Social Network Analysis Toolbox in Networking



- Computer networks : systems of increasing complexity

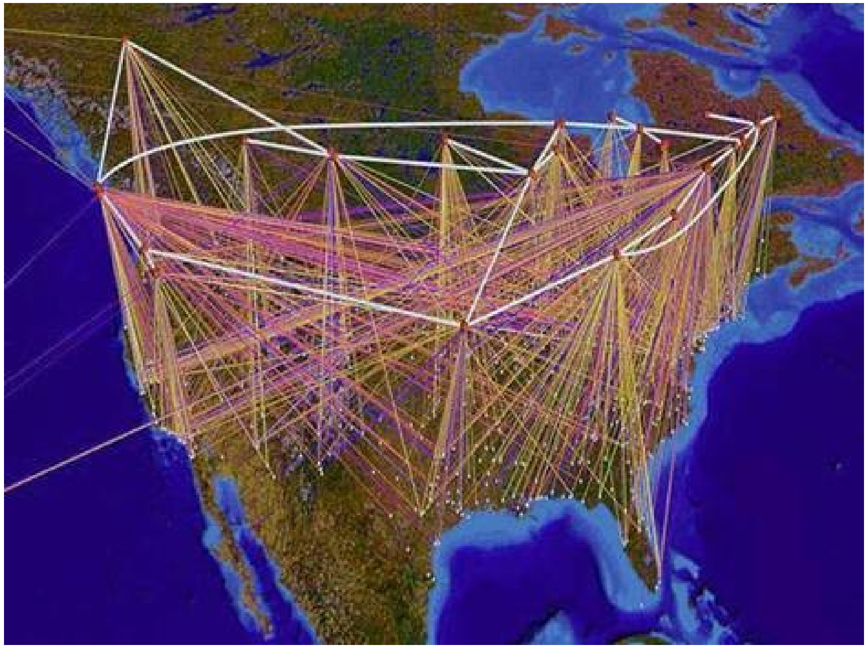
Social Network Analysis Toolbox in Networking



- Computer networks : systems of increasing complexity
- Difficult to design efficient network protocols

HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

Social Network Analysis Toolbox in Networking



- Computer networks : systems of increasing complexity
- Difficult to design efficient network protocols
- SNA: Analytical framework for understanding structural properties

The importance of being.. central

- Tasks of service placement, data caching, content forwarding..

The importance of being.. central

- Tasks of service placement, data caching, content forwarding..
- A number of relevant protocol instances seek to identify the central one(s)!
- Usually the ranking of the metric values

matters, rather than the absolute values

The importance of being.. central

- Tasks of service placement, data caching, content forwarding..
- A number of relevant protocol instances seek to identify the central one(s)!
- Usually the ranking of the metric values

matters, rather than the absolute values

- Centrality: a measure of

importance (sociological origin)

- Different Interpretations

related to the way traffic flows

  - Betweenness Centrality (BC)

Centrality computations under different scope

- Computations require global topological info
  - Problematic in large scale networks
  - Infeasible in self-organizing environments

Centrality computations under different scope

- Computations require global topological info
  - Problematic in large scale networks
  - Infeasible in self-organizing environments

- A realistic alternative: limit the computations in the ego-network
  - A subgraph involving the reference node (ego),
  - its 1-hop neighbors,
  - and their interconnection

Centrality computations under different scope

- Computations require global topological info
  - Problematic in large scale networks
  - Infeasible in self-organizing environments

Ego-network
$G^u_1=(V^u_1,E^u_1)$

- A realistic alternative: limit the computations in the ego-network
  - A subgraph involving the reference node (ego),
  - its 1-hop neighbors,
  - and their interconnection

Main research question

- How well do these local metrics approximate the real ones?

- Networking community seems to take it for granted!
  - The use of ego-metrics is based on (rank)-correlation values of 0.9
  - The studied network topologies are in many cases not relevant
  - Content-related protocols are likely to operate over router-level topologies
- Does high correlation imply efficiency of protocols that employ ego-metrics?

The studied centrality indices

	Socio-centric	Ego-centric
Betweenness Centrality (BC)	$BC(u) = \sum_{\substack{s,t \in V \\ s<t}} \frac{\sigma_{st}(u)}{\sigma_{st}}$	$egoBC(u;r) = BC(u)_{\mid V=V_r^u}$
Conditional Betweenness Centrality (CBC)	$CBC(u;t) = \sum_{\substack{s \in V \\ s \neq t}} \frac{\sigma_{st}(u)}{\sigma_{st}}$ $\sigma_{st}(s) = 0$	$egoCBC(u;t,r) = \sum_{\substack{s \in V_r^u \\ t' \in e_r(u;t)}} \frac{\sigma_{st'}(u)}{\sigma_{st'}} \mathbf{1}_{\{h(s,t') \leq h(s,l),\ l \in e_r(u;t)\}}$
Degree Centrality (DC)	Number of 1-hop neighbors	

The studied centrality indices

a measure of the importance
of node's u social position :
lies on paths linking others

	Socio-centric	Ego-centric
Betweenness Centrality (BC)	$BC(u) = \sum_{\substack{s,t \in V \\ s<t}} \frac{\sigma_{st}(u)}{\sigma_{st}}$	$egoBC(u;r) = BC(u)_{\mid V=V_r^u}$
Conditional Betweenness Centrality (CBC)	$CBC(u;t) = \sum_{\substack{s \in V \\ s \neq t}} \frac{\sigma_{st}(u)}{\sigma_{st}}$ $\sigma_{st}(s) = 0$	$egoCBC(u;t,r) = \sum_{\substack{s \in V_r^u \\ t' \in e_r(u;t)}} \frac{\sigma_{st'}(u)}{\sigma_{st'}} \mathbf{1}_{\{h(s,t') \leq h(s,l),\ l \in e_r(u;t)\}}$
Degree Centrality (DC)	Number of 1-hop neighbors	

The studied centrality indices

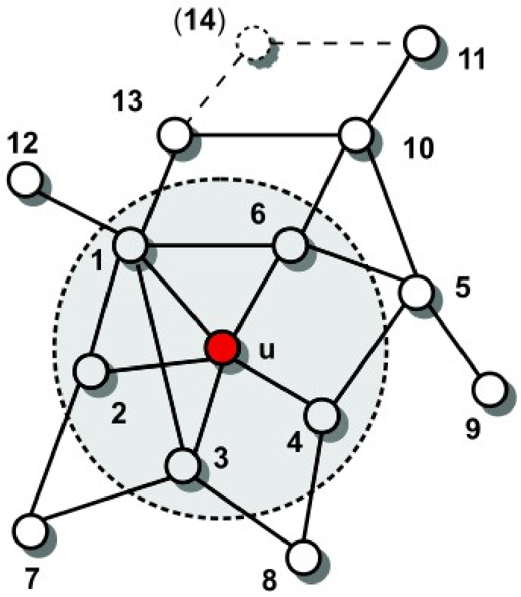
a measure of the importance
of node's u social position :
lies on paths linking others

	Socio-centric	Ego-centric
Betweenness Centrality (BC)	$BC(u) = \sum_{\substack{s,t\in V\\ s<t}} \frac{\sigma_{st}(u)}{\sigma_{st}}$	$egoBC(u;r) = BC(u)_{\vert V=V_r^u}$
Conditional Betweenness Centrality (CBC)	$CBC(u;t) = \sum_{\substack{s\in V\\ s\neq t}} \frac{\sigma_{st}(u)}{\sigma_{st}}$ $\sigma_{st}(s) = 0$	$egoCBC(u;t,r) = \sum_{\substack{s\in V_r^u\\ t'\in e_r(u;t)}} \frac{\sigma_{st'}(u)}{\sigma_{st'}} \mathbf{1}_{\{h(s,t')\leq h(s,l),\ l\in e_r(u;t)\}}$
Degree Centrality (DC)	Number of 1-hop neighbors	

a measure of the importance
of node's u social position : ability to
control information flow towards
target node

The studied centrality indices

- Egocentric conditional BC



- For a given destination t

identify the set of *exit* nodes $e_r(u;t)$

- $e_1(u;11) = \{6\}$
- $e_1(u;9) = \{4,6\}$

- Nodes 2, 3 and 4 contribute to egoCBC(u;11,1) = 2

with contributions 1/2, 1/2 and1, respectively.

Computational benefits of local metrics

- D: network diameter
- d_{max}: maximun degree

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

- d_{max} typically smaller than |V|

Background in (geodesic) centrality computations

Scope of the input (topological) info?

Background in (geodesic) centrality computations

Scope of the input (topological) info?

Local information

Is there positive correlation between local and global metrics?

- BC vs egoBC for small social nets and random graphs (Marsden, Borgatti)

-linear BC-DC relationship in AS maps (Vázquez, et al.)

-localized bridging centrality and volume centrality correlate with bridging and closeness centrality, respectively

Background in (geodesic) centrality computations

Scope of the input (topological) info?

Local information

Is there positive correlation between local and global metrics?

- BC vs egoBC for small social nets and random graphs (Marsden, Borgatti)

-linear BC-DC relationship in AS maps (Vázquez, et al.)

-localized bridging centrality and volume centrality correlate with bridging and closeness centrality, respectively

Given that there is positive correlation, how can I exploit it?

- identify users with rich social nets in large collaborative nets (Daly)
-BubbleRap DTN forwarding (Hui et al.)
-Selective Caching in CCNs (Chai et al.)

Background in (geodesic) centrality computations

Scope of the input (topological) info?

Local information

Is there positive correlation between local and global metrics?

- BC vs egoBC for small social nets and random graphs (Marsden, Borgatti)

-linear BC-DC relationship in AS maps (Vázquez, et al.)

-localized bridging centrality and volume centrality correlate with bridging and closeness centrality, respectively

Given that there is positive correlation, how can I exploit it?

- identify users with rich social nets in large collaborative nets (Daly)
-BubbleRap DTN forwarding (Hui et al.)
-Selective Caching in CCNs (Chai et al.)

Global information

Exact algorithms

Brandes

Approximation algorithms

Pivot-BC,
Scale-BC,
k-BC

Background in (geodesic) centrality computations

Scope of the input (topological) info?

Local information

Is there positive correlation between local and global metrics?

- BC vs egoBC for small social nets and random graphs (Marsden, Borgatti)

-linear BC-DC relationship in AS maps (Vázquez, et al.)

-localized bridging centrality and volume centrality correlate with bridging and closeness centrality, respectively

Given that there is positive correlation, how can I exploit it?

- identify users with rich social nets in large collaborative nets (Daly)
-BubbleRap DTN forwarding (Hui et al.)
-Selective Caching in CCNs (Chai et al.)

Distributed computation (locally determined based on global info)

RW- schemes
RW-betweenness,
2^{nd} order centrality (Kermarrec et al.)
Rw-sampling (Lim-Towsley)

They require info gathering from the whole or part of the network topology

Global information

Exact algorithms

Brandes

Approximation algorithms

Pivot-BC,
Scale-BC,
k-BC

Background in (geodesic) centrality computations

Scope of the input (topological) info?

Local information

Is there positive correlation between local and global metrics?

- BC vs egoBC for small social nets and random graphs (Marsden, Borgatti)

-linear BC-DC relationship in AS maps (Vázquez, et al.)

-localized bridging centrality and volume centrality correlate with bridging and closeness centrality, respectively

Given that there is positive correlation, how can I exploit it?

- identify users with rich social nets in large collaborative nets (Daly)
-BubbleRap DTN forwarding (Hui et al.)
-Selective Caching in CCNs (Chai et al.)

Distributed computation (locally determined based on global info)

RW- schemes
RW-betweenness,
2nd order centrality (Kermarrec et al.)
Rw-sampling (Lim-Towsley)

They require info gathering from the whole or part of the network topology

Global information

Exact algorithms

Brandes

Approximation algorithms

Pivot-BC,
Scale-BC,
k-BC

Background in (geodesic) centrality computations

Scope of the input (topological) info?

Local information

Is there positive correlation between local and global metrics?

- BC vs egoBC for small social nets and random graphs (Marsden, Borgatti)

-linear BC-DC relationship in AS maps (Vázquez, et al.)

-localized bridging centrality and volume centrality correlate with bridging and closeness centrality, respectively

Given that there is positive correlation, how can I exploit it?

- identify users with rich social nets in large collaborative nets (Daly)
-BubbleRap DTN forwarding (Hui et al.)
-Selective Caching in CCNs (Chai et al.)

Distributed computation (locally determined based on global info)

RW- schemes
RW-betweenness,
2nd order centrality (Kermarrec et al.)
Rw-sampling (Lim-Towsley)

They require info gathering from the whole or part of the network topology

Global information

Exact algorithms

Brandes

Approximation algorithms

Pivot-BC,
Scale-BC,
k-BC

What about router-level topologies of thousands of nodes?

Background in (geodesic) centrality computations

Scope of the input (topological) info?

Local information

Is there positive correlation between local and global metrics?

- BC vs egoBC for small social nets and random graphs (Marsden, Borgatti)

-linear BC-DC relationship in AS maps (Vázquez, et al.)

-localized bridging centrality and volume centrality correlate with bridging and closeness centrality, respectively

Given that there is positive correlation, how can I exploit it?

- identify users with rich social nets in large collaborative nets (Daly)
-BubbleRap DTN forwarding (Hui et al.)
-Selective Caching in CCNs (Chai et al.)

Distributed computation (locally determined based on global info)

RW- schemes
RW-betweenness,
2nd order centrality (Kermarrec et al.)
Rw-sampling (Lim-Towsley)

They require info gathering from the whole or part of the network topology

Global information

Exact algorithms

Brandes

Approximation algorithms

Pivot-BC,
Scale-BC,
k-BC

What about router-level topologies of thousands of nodes?

Is any positive correlation enough to guarantee efficiency for the local-info-based protocol instances ?

Capturing correlation (between node rankings)

- Spearman correlation coefficient

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

  - ranks of each graph node when ordered according to the sociocentric and egocentric definition of the metrics

- Top-k overlap
  - Overlap between the k nodes exhibiting the top values of each metric

- Pearson correlation coefficient

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

  - Pairs of the socio and ego-betweenness variants (sB(u),eB(u)) of each node u

Internet router-level topologies

- `mrinfo` topologies
  - 14 different AS topologies (Tier-1 and Transit ISPs)
  - Collected during 2004-2008
  - Multicast discovering tool
- `Rocketfuel` topologies
  - Widely used in experimental studies
  - 800 vantage points serving as `traceroute` sources
  - Innovative techniques to address the alias-resolution problem
- `Caida` topologies
  - Collected during Oct.-Nov. 2011
  - `Traceroute` probes to randomly chosen destinations from 54 monitors worldwide
  - Aim was to discover the largest ISP networks present in the dataset

Correlation insights from a synthetic topology

- BC - egoBC correlation on a rectangular grid
- Fixed ego-network sizes
- EgoBC index attains three values (r=1)
- Rank correlation

decreases with grid size

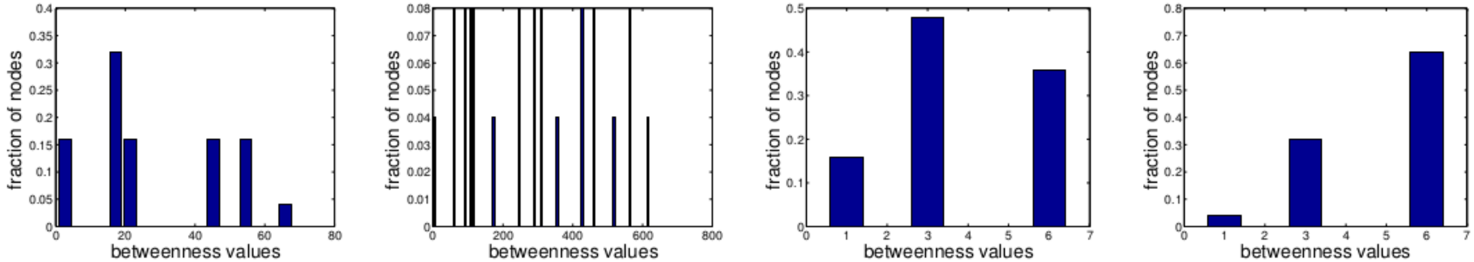
Grid size	Diameter / Mean degree	Spearman ρ ego-network (**r=1**)	Spearman ρ ego-network (**r=2**)
5x5	8 / 3.200	0.9195	0.9679
10x10	18 / 3.600	0.8400	0.9556
20x20	38 / 3.800	0.6802	0.8459
50x50	98 / 3.920	0.2429	0.2942
60x8	66 / 3.717	0.5735	0.6336
90x8	96 / 3.728	0.5390	0.5870
150x8	156 / 3.737	0.4584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

Correlation insights from a synthetic topology

- BC - egoBC correlation on a rectangular grid
- Fixed ego-network sizes
- EgoBC index attains three values (r=1)
- Rank correlation

decreases with grid size

Grid size	Diameter / Mean degree	Spearman ρ ego-network (**r=1**)	Spearman ρ 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



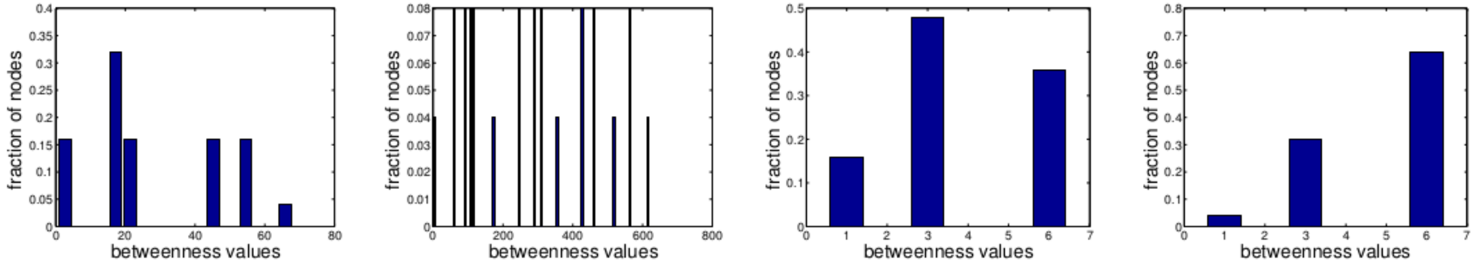
a. BC (5x5) b. BC (10x10) c. egoBC (5x5) d. egoBC (10x10)

Correlation insights from a synthetic topology

- BC - egoBC correlation on a rectangular grid
- Fixed ego-network sizes
- EgoBC index attains three values (r=1)
- Rank correlation

decreases with grid size

Grid size	Diameter / Mean degree	Spearman ρ ego-network (**r=1**)	Spearman ρ 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)

- Size scaling: the BC spectrum is getting richer!

Correlation is high in real-world ISP topologies

- High positive correlation (0.8-0.9) between BC - egoBC for both Spearman and Pearson already with r=1
- “Asymmetry” that yields a wide range of BC and egoBC values

DataSet	ID	ISP(AS number)	<CC>	Diameter	Size	<degree>	**BC** *vs.* **ego-BC**			
							Spearman ρ		Pearson r_{Prs}	
							ego-net. **r=1**	ego-net. **r=2**	ego-net. **r=1**	ego-net. **r=2**
R	61	Ebone(1755)	0.115	13	295	3.68	0.9736	0.9860	0.6856	0.8895
O	62	Tiscali(3257)	0.028	14	411	3.18	0.9522	0.9659	0.6073	0.9281
C	63	Exodus(3967)	0.273	14	353	4.65	0.9125	0.9792	0.6100	0.9061
K	64	Telstra (1221)	0.015	15	2515	2.42	0.9990	0.9990	0.3336	0.7565
E	65	Sprint(1239)	0.022	13	7303	2.71	0.9980	0.9990	0.4770	0.7977
T	66	Level-3(3356)	0.097	10	1620	8.32	0.9841	0.9923	0.6346	0.9075
F	67	AT&T(7018)	0.005	14	9418	2.48	0.9988	0.9994	0.3388	0.5302
L	68	Verio (2914)	0.071	15	4607	3.28	0.9904	0.9969	0.4729	0.8044

Correlation is high in real-world ISP topologies

- High positive correlation (0.8-0.9) between BC - egoBC for both Spearman and Pearson already with r=1
- “Asymmetry” that yields a wide range of BC and egoBC values

DataSet	ID	ISP(AS number)	<CC>	Diameter	Size	<degree>	BC *vs.* ego-BC			
							Spearman ρ		Pearson r_{Prs}	
							ego-net. **r=1**	ego-net. **r=2**	ego-net. **r=1**	ego-net. **r=2**
	70	UUNet (701)	0.012	15	18281	2.77	0.9841	0.9886	0.5430	0.8752
C	71	COGENT/PSI(174)	0.062	32	14413	3.09	0.9638	0.9599	0.7272	0.9354
A	72	LDComNet(15557)	0.021	40	6598	2.47	0.9674	0.9245	0.3782	0.7676
I	74	ChinaTelecom(4134)	0.083	19	81121	3.97	0.8324	0.8986	0.7861	0.9714
D	75	FUSE-NET(6181)	0.018	10	1831	2.38	0.9903	0.9763	0.6205	0.8574
A	76	JanetUK(786)	0.031	24	2259	2.26	0.9819	0.9834	0.4444	0.8506

Correlation is high in real-world ISP topologies

- High positive correlation (0.8-0.9) between BC - egoBC for both Spearman and Pearson already with r=1
- “Asymmetry” that yields a wide range of BC and egoBC values

DataSet	ID	ISP(AS number)	<CC>	Diameter	Size	<degree>	**BC** *vs.* **ego-BC** Spearman ρ ego-net. **r=1**	Spearman ρ ego-net. **r=2**	Pearson r_{Prs} ego-net. **r=1**	Pearson r_{Prs} ego-net. **r=2**
	35	Global Crossing(3549)	0.479	9	100	3.78	0.9690	0.9853	0.7029	0.9255
m	33	NTTC-Gin(2914)	0.307	11	180	3.53	0.9209	0.9565	0.7479	0.8561
r	13	Level-3(3356)	0.169	25	378	4.49	0.2708	0.9393	-0.0918	0.7982
i	12	-//-	0.149	28	436	4.98	0.2055	0.9381	-0.1217	0.7392
n	20	Sprint(1239)	0.287	16	528	3.13	0.9866	0.9928	0.5805	0.8488
f	38	Iunet(1267)	0.231	12	645	3.75	0.8790	0.9516	0.9094	0.9568
o	44	Telecom Italia(3269)	0.037	13	995	3.65	0.7950	0.9828	0.3362	0.8699
	50	TeleDanmark(3292)	0.058	15	1240	3.06	0.9569	0.9738	0.5475	0.9025

Correlation is high in real-world ISP topologies

- High positive correlation (0.8-0.9) between BC - egoBC for both Spearman and Pearson already with r=1
- “Asymmetry” that yields a wide range of BC and egoBC values
- A notable exception

DataSet	ID	ISP(AS number)	$<CC>$	Diameter	Size	<degree>	**BC** *vs.* **ego-BC**			
							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
m	33	NTTC-Gin(2914)	0.307	11	180	3.53	0.9209	0.9565	0.7479	0.8561
r	13	Level-3(3356)	0.169	25	378	4.49	0.2708	0.9393	-0.0918	0.7982
i	12	-//-	0.149	28	436	4.98	0.2055	0.9381	-0.1217	0.7392
n	20	Sprint(1239)	0.287	16	528	3.13	0.9866	0.9928	0.5805	0.8488
f	38	Iunet(1267)	0.231	12	645	3.75	0.8790	0.9516	0.9094	0.9568
o	44	Telecom Italia(3269)	0.037	13	995	3.65	0.7950	0.9828	0.3362	0.8699
	50	TeleDanmark(3292)	0.058	15	1240	3.06	0.9569	0.9738	0.5475	0.9025

HELLENIC REPUBLIC
National and Kapodistrian
University of Athens

One outlier ISP topology

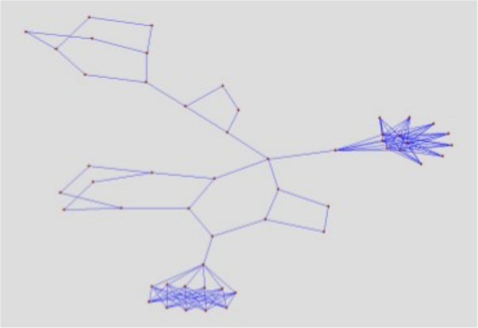
- Pathologies in the `mrinfo` Level-3 ISP snapshots

One outlier ISP topology

- Pathologies in the `mrinfo` Level-3 ISP snapshots
  - The only one with very low Spearman and negative Pearson correlation values out of 21 different ISP networks

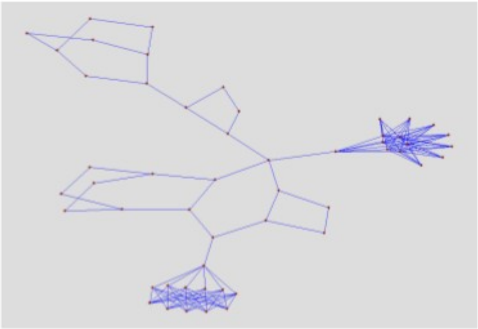
One outlier ISP topology

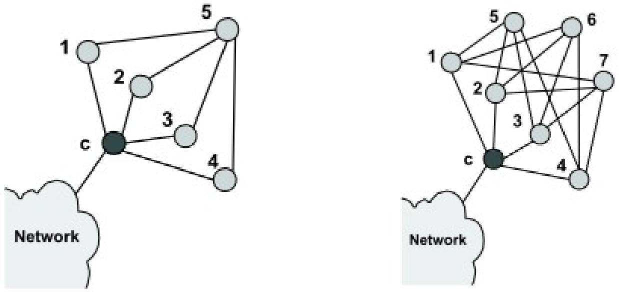
- Pathologies in the `mrinfo` Level-3 ISP snapshots
  - The only one with very low Spearman and negative Pearson correlation values out of 21 different ISP networks



One outlier ISP topology

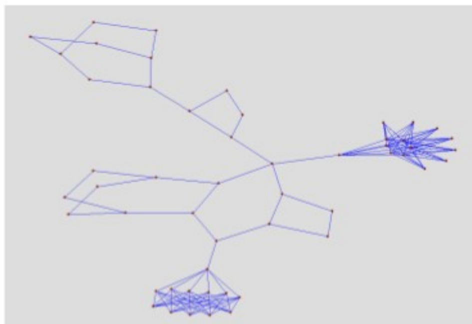
- Pathologies in the `mrinfo` Level-3 ISP snapshots
  - The only one with very low Spearman and negative Pearson correlation values out of 21 different ISP networks

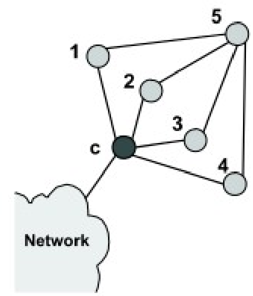


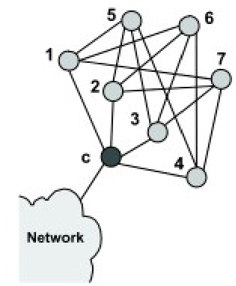


One outlier ISP topology

- Pathologies in the `mrinfo` Level-3 ISP snapshots
  - The only one with very low Spearman and negative Pearson correlation values out of 21 different ISP networks



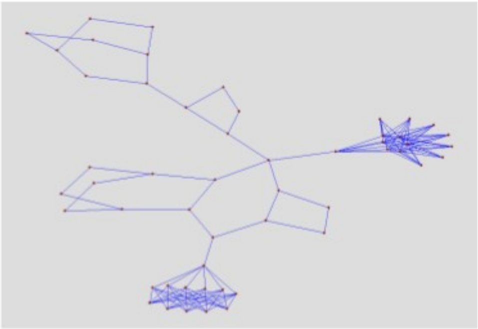


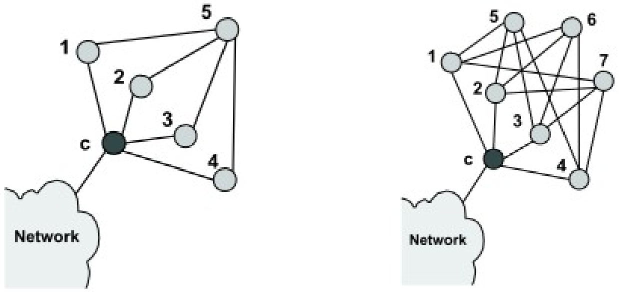


- Clustered structures of nodes that exhibit higher egoBC than global BC values

One outlier ISP topology

- Pathologies in the `mrinfo` Level-3 ISP snapshots
  - The only one with very low Spearman and negative Pearson correlation values out of 21 different ISP networks





- Clustered structures of nodes that exhibit higher egoBC than global BC values
- More general result : the actual association between the local and global metrics is not determined solely by the degree distribution

Correlation is high in real-world ISP topologies

- Correlation between BC and degree (DC)
- High Pearson and even higher Spearman
- Consistently lower than the BC- egoBC
- Extends the previously reported BC-DC correlation measured over AS-level topologies to the router-level ones

DataSet	ID	ISP(AS number)	**BC *vs.* DC**	
			Spearman ρ	Pearson r_{Prs}
	35	Global Crossing(3549)	0.8506	0.6714
m	33	NTTC-Gin(2914)	0.8180	0.6664
r	13	Level-3(3356)	0.1953	-0.0813
i	12	-//-	0.1696	-0.1128
n	20	Sprint(1239)	0.8543	0.6815
f	38	Iunet(1267)	0.8549	0.7708
o	44	Telecom Italia(3269)	0.7733	0.4852
	50	TeleDanmark(3292)	0.9388	0.5538
R	61	Ebone(1755)	0.9443	0.7457
O	62	Tiscali(3257)	0.9464	0.7103
C	63	Exodus(3967)	0.8204	0.6241
K	64	Telstra (1221)	0.9783	0.5172
E	65	Sprint(1239)	0.9562	0.6537
T	66	Level-3(3356)	0.9655	0.7045
F	67	AT&T(7018)	0.9882	0.4483
L	68	Verio (2914)	0.9315	0.6718
	70	UUNet (701)	0.9694	0.7544
C	71	COGENT/PSI(174)	0.8940	0.8791
A	72	LDComNet(15557)	0.9479	0.6634
I	74	ChinaTelecom(4134)	0.7370	0.8795
D	75	FUSE-NET(6181)	0.9536	0.7445
A	76	JanetUK(786)	0.9450	0.5765

Correlation is high in real-world ISP topologies

- Correlation between the conditional BC variants (CBC-egoCBC)

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

- Considerably increased coefficient values for the Level-3 outliers
- egoCBC(u;t;r) considers only the paths that lead to the target t
- The differences between the two counterparts may occur a certain angle that encompasses t

..but this is not uniform over the full ranking

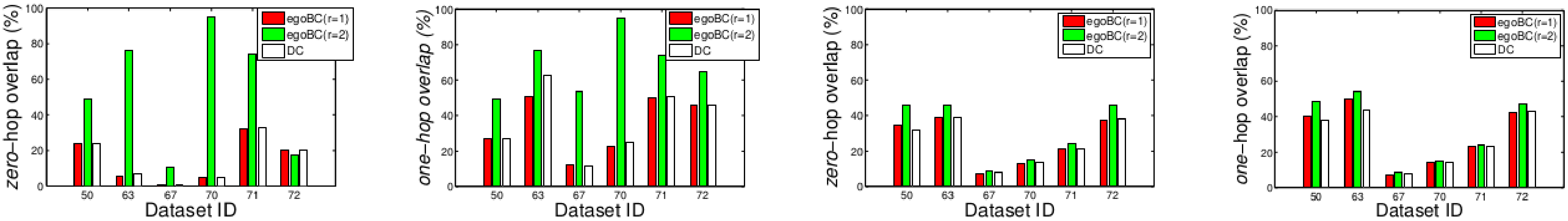
- Overlap between nodes with the top-*k* nodes local centrality and BC values

	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

- Low overlap values do not contradict our previous results
- The previously observed high correlation is due to nodes of lower rank
  - Nodes with zero egoBC and BC values have been reported in literature to drastically contribute to high correlation values
- A warning sign regarding what high correlation can reveal about the practical implications of local centrality metrics

Practical utility of local centrality metrics (1/3)

- Motivation: a high-degree file-searching scheme in unstructured peer-to-peer has been proved more efficient random walks
- Implement a local centrality-driven navigation scheme of MAX or MAX-MIN pattern with respect to node centrality
- Random selection of starting points
- α-hops overlap measures the percentage of the final locations lying within α hops away from those the global metric yields



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

- Overlap between the final locations achieved with local and global centrality metrics as driver

Practical utility of local centrality metrics (2/3)

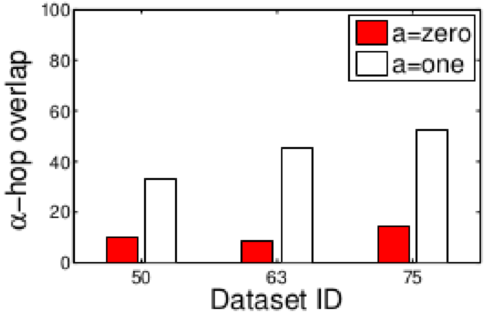
- How does the crawler navigate the network?
- Driven by local metrics it consistently takes, up to 2.3 on average, less hops than when it is BC-driven
- It fails to identify the same sequence of central nodes
- Numerical summaries the coefficients provide, fail to capture in micro-level the relative significance

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

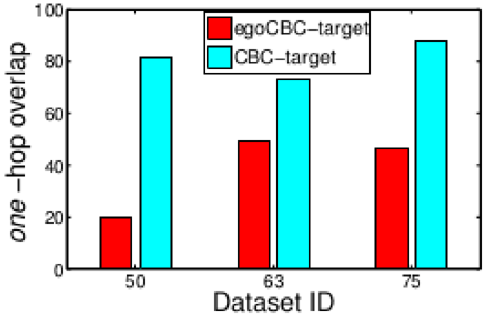
Mean hopcount and unique final locations (UFL) for the MAX pattern centrality-driven navigation scheme

Practical utility of local centrality metrics (3/3)

- Conditional centrality metrics involve a target node
- We are enabled to compare CBC and egoCBC over a search scheme



a. MAX search pattern



b. MAX search pattern

- The number of targets reached by the local CBC variant is in good agreement with the discovered P2P nodes [Adamic], using DC as a driver
- Low overlap between the final locations achieved by the two counterparts
- The hopcount to the final location is again measured consistently lower (i.e., 0.3 to 1.5 hops) for the egoCBC case

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

Take-home remarks

- Can we substitute the original centrality metrics with their computationally friendly local approximations in ISP topologies?

- In terms of correlation...
  - High rank correlation measured across all (20) datasets but one
  - The match between the top-k nodes selected by local and global centrality is found low

- So, what can the high values of the correlation coefficients reveal about the performance of network functions?

  - Simple navigation/search schemes employing local centrality metrics produce significantly different navigation patterns/lower hit-rates than the original global metrics do

Take-home remarks

- Can we substitute the original centrality metrics with their computationally friendly local approximations in ISP topologies?

- In terms of correlation...
  - High rank correlation measured across all (20) datasets but one
  - The match between the top-k nodes selected by local and global centrality is found low

- So, what can the high values of the correlation coefficients reveal about the performance of network functions?

  - Simple navigation/search schemes employing local centrality metrics produce significantly different navigation patterns/lower hit-rates than the original global metrics do

We warn against relying on the correlation indices for assessing the substitutability of ego- and sociocentric variants of centrality metrics

Thank you!

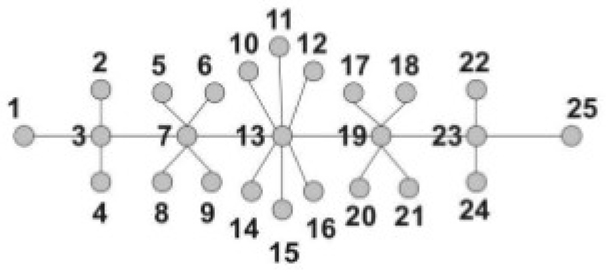
Thank you!

Questions ?

Back-up slides

Centrality-based navigation patterns are different despite the positive egoBC-BC correlation

- Toy topology with perfect rank- correlation



- Removing sequentially the nodes 5,6,17 and 18,
- Rank-correlation reduces from 1 to 0.9953
- Zero-hop overlap for the MAX pattern diminishes from 100% to 61.90%.
- The numerical summaries provided by the coefficients fail to reflect in micro-level the relative significance of each node

EgoBC-BC “micro”-correlation

- For every node identify the set of its first neighbors
- Compute the Spearman coefficient for EgoBC and BC values of the set

- Dataset 50
  - ρ>0.85 : 51.45%
  - ρ<0 : 4.83%
  - P=-1 : 3.47%