Exploiting Social Metrics in Content Distribution

Ioannis Stavrakakis

National & Kapodistrian University of Athens

Based on works with:

Merkouris Karaliopoulos, Eva Jaho, Panagiotis Pandazopoulos, Pavlos Nikolopoulos, Therapon Papadimitriou

June 30, 2011

Focus on two key social metrics

- ❑ Interest similarity
- ❑ Centrality

1 - Interest Similarity

- Groups in online social networks formed to exchange information / content based on acquaintance, family relationships, social status, educational/professional background, ...

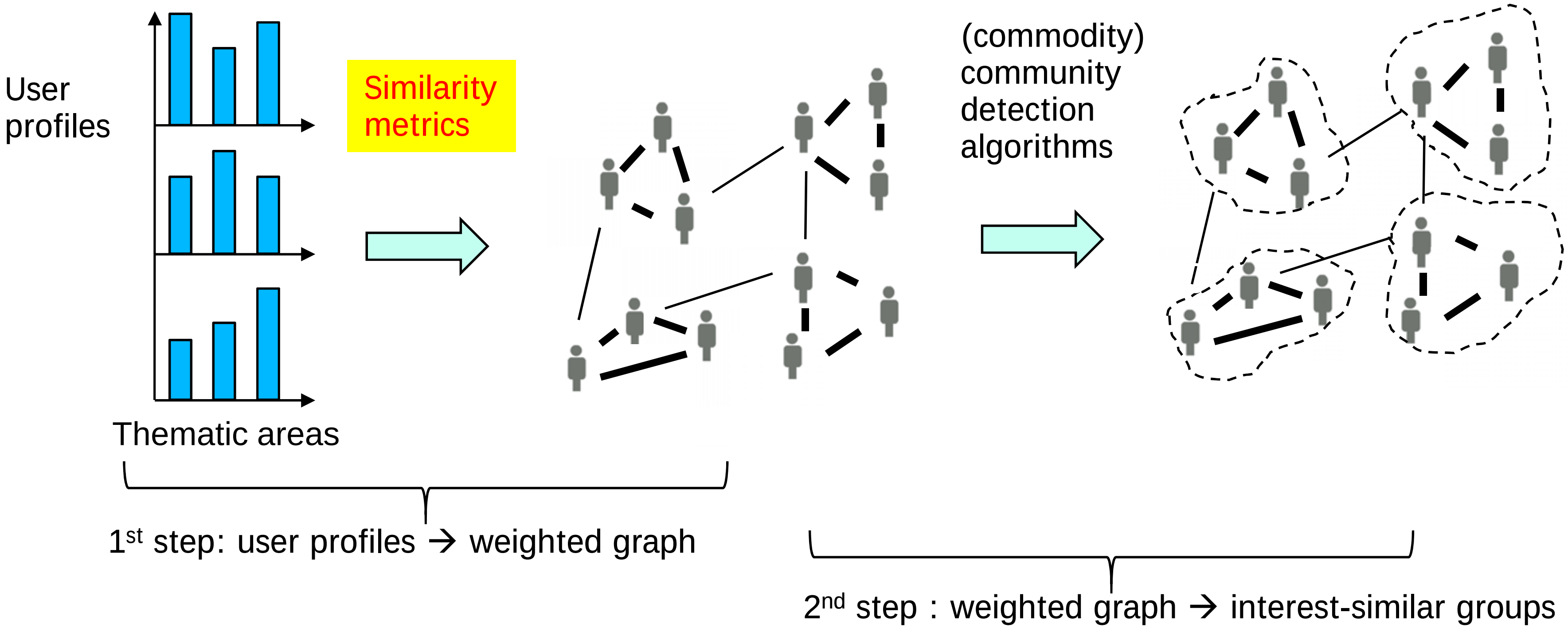
- ...*yet* ***interests/preferences*** of group members are not always similar

Is there value in assessing and exploiting interest similarity in groups?

Define and measure Interest Similarity

- ❑ assess overall interest similarity in social groups
- ❑ identify interest-based structure within social groups

ISCoDE framework[3]



[3]E. Jaho, M. Karaliopoulos, I. Stavrakakis. ISCoDe: a framework for interest similarity-based community detection in social networks. Third International Workshop on Network Science for Communication Networks (INFOCOM-NetSciCom'11), Apr. 10-15, 2011, Shanghai.

Similarity metrics: PS vs. InvKL (Kullback Leibler)

- Proportional Similarity (PS)
  - PS : {F^i , F^j} → [0,1]

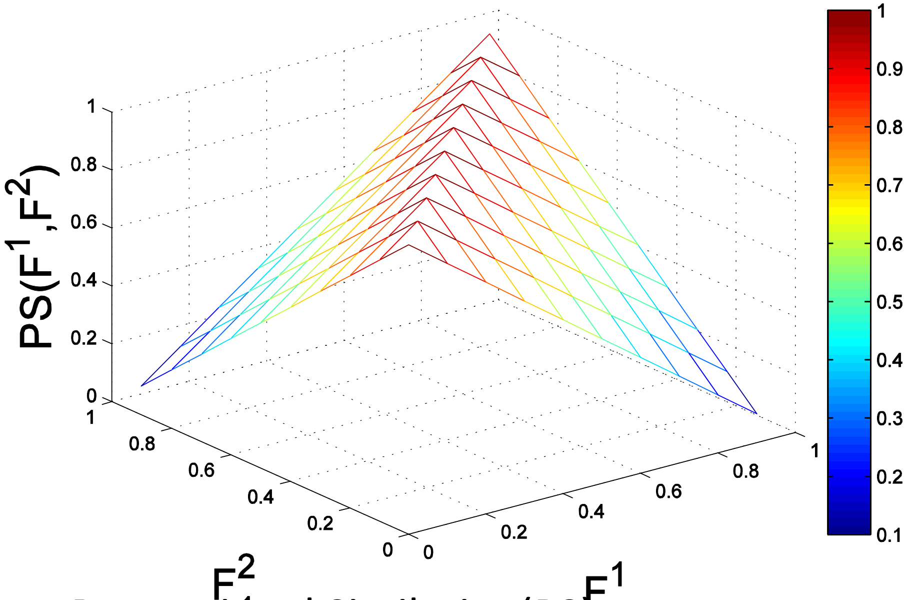
$$PS(F^i,F^j)=1-\frac{1}{2}\sum_{m=1}^{M}\left|F^i{}_m-F^j{}_m\right|$$

- Inverse symmetrized KL divergence
  - InvKL : {F^i, F^j} → (0,∞)

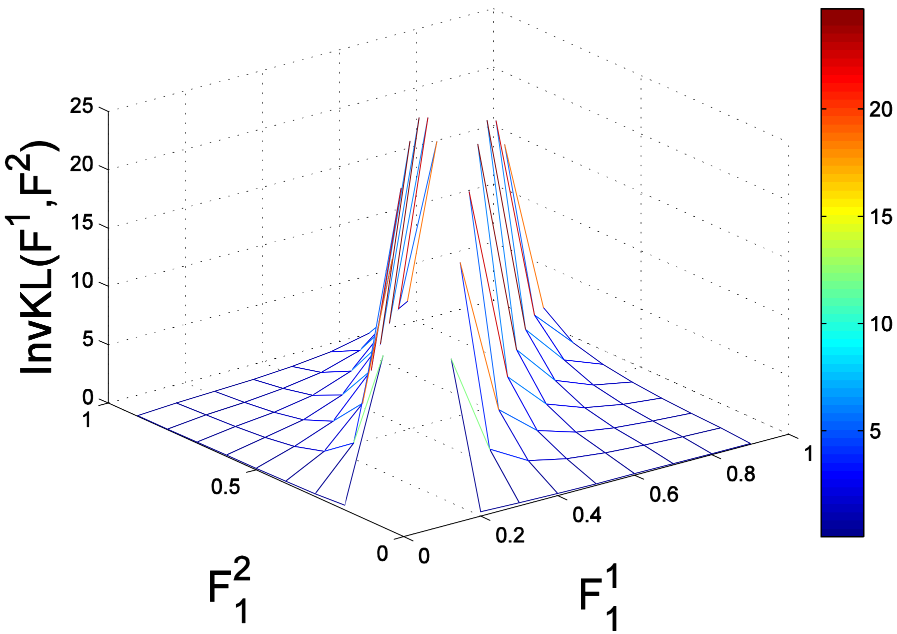
$$InvKL(F^i,F^j)=\frac{1}{\sum_{m=1}^{M}F^i{}_m\log\frac{F^i{}_m}{F^j{}_m}+\sum_{m=1}^{M}F^j{}_m\log\frac{F^j{}_m}{F^i{}_m}}$$

$F^n{}_m$, 1≤n≤N, 1≤m≤M : distribution of node n over interest class m

Example with M=2 interest classes and N=2 nodes



- Proportional Similarity (PS)



- Inverse symmetrized KL divergence (InvKL)

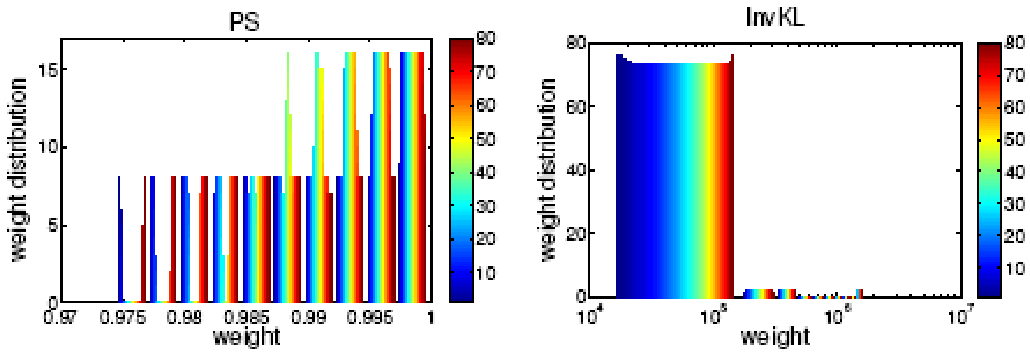
Resolution performance

(a) Similar nodes

PS			InvKL		
Q	C	partition	Q	C	partition
0.0215	2	{1..38} {39..80}	0.6740	5	{1..14} {15..28} {29..44} {45..61} {62..80}

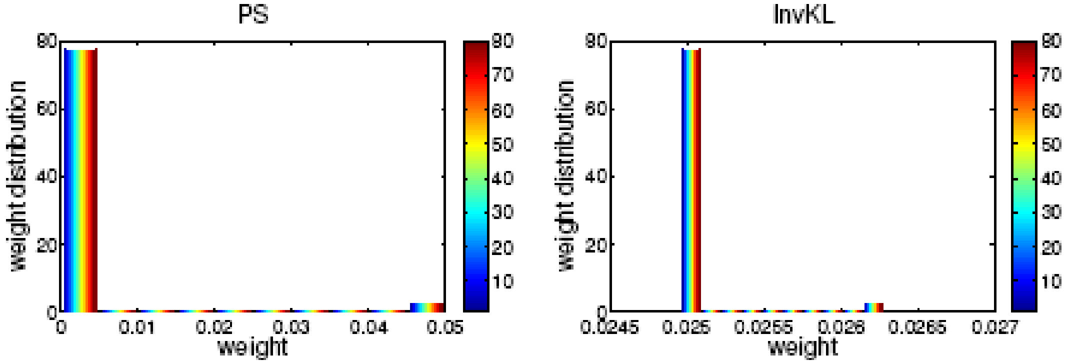
(b) Dissimilar nodes

PS			InvKL		
Q	C	partition	Q	C	partition
0.7860	10	{1..8}..{73..80}	0	1	{1..80}



(a) Similar nodes

InvKL can identify smaller and more similar communities than PS, in a highly similar network



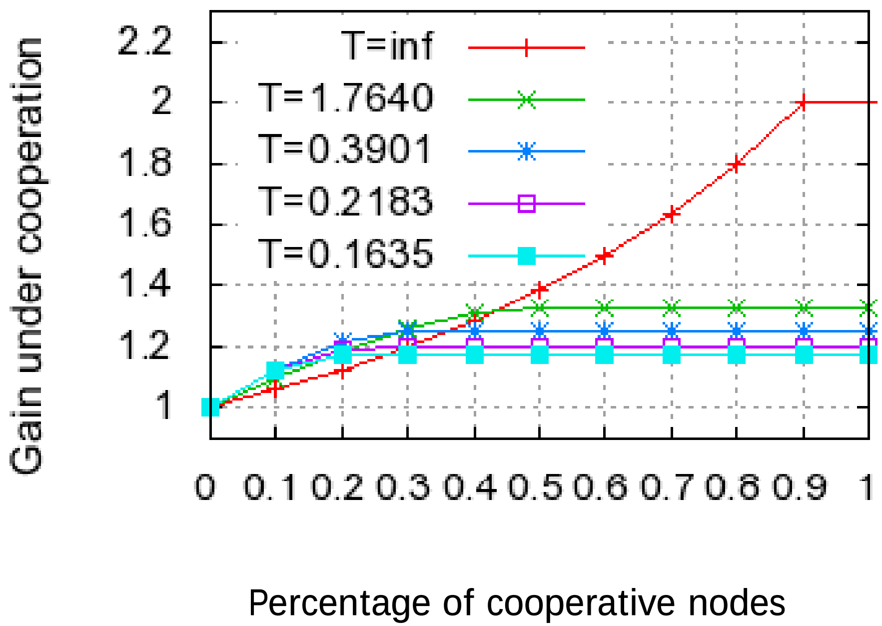
(b) Dissimilar nodes

PS can identify smaller communities than InvKL, in a highly dissimilar network

(could argue that this is not very useful)

Can Interest similarity improve network protocols ?

- Gain of cooperation for content replication in a group of nodes[1]

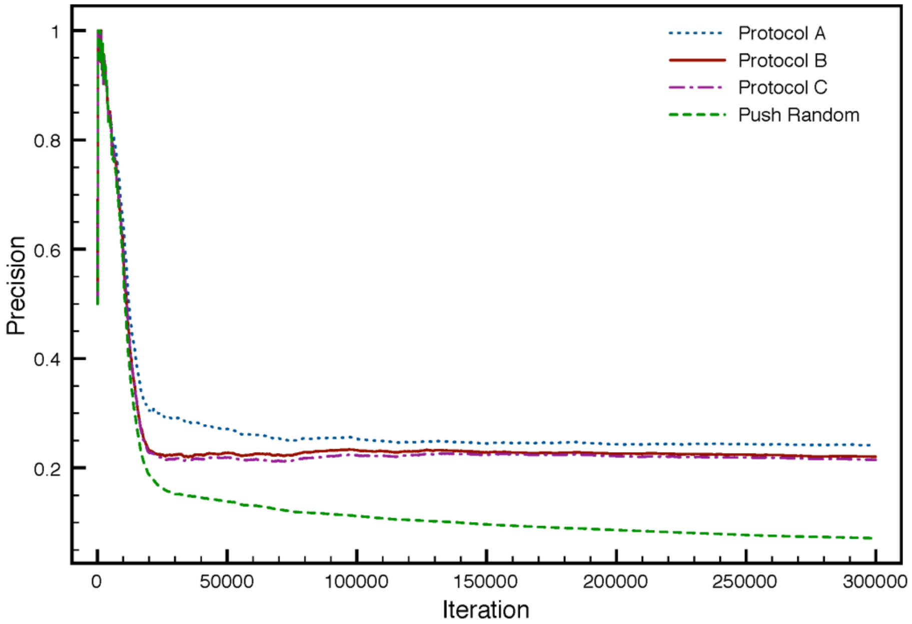


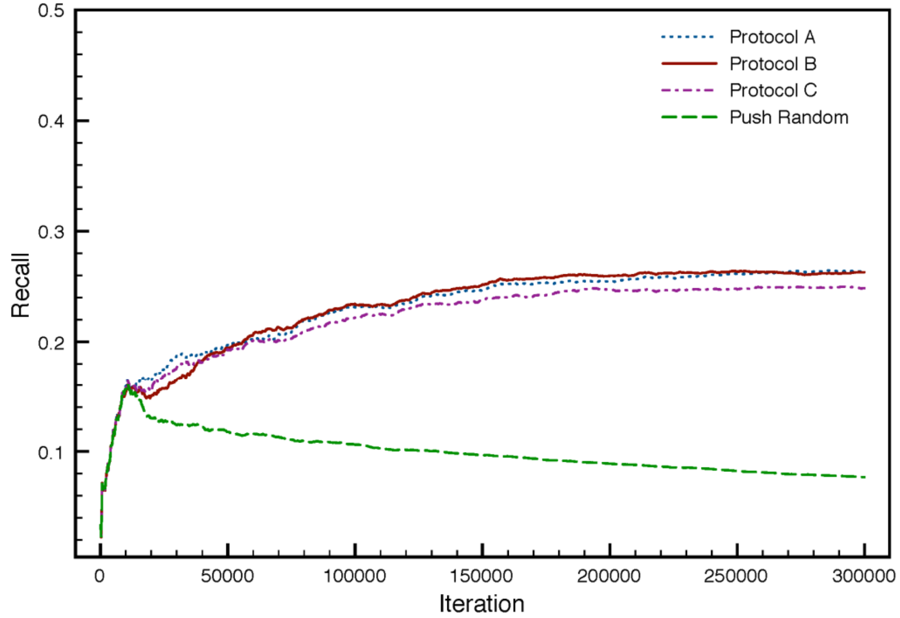
- T : tightness metric (=mean invKL), measuring interest similarity across group members

[1] E. Jaho, M. Karaliopoulos, I. Stavrakakis, “Social similarity as a driver for selfish, cooperative and altruistic behavior”, in Proc. AOC 2010 (extended version submitted to IEEE TPDS)

Can Interest similarity improve network protocols ?

- Content dissemination in opportunistic networks[2]





- Protocols A,B,C are push protocols exercising *interest-based* forwarding

(Precision: valuable received / total received --– Recall: valuable received / total potentially valuable generated)

[2]S.M. Allen, M.J. Chorley, G.B. Colombo, E. Jaho, M. Karaliopoulos, I. Stavrakakis, R.M Whitaker, "Exploiting user interest similarity and social links for microblog forwarding in mobile opportunistic networks", submitted to Elsevier PMC, 2011

2- Betweeness Centrality (BC)

2.1 - Content (service) Migration / Placement

Can BC help provide for a low-complexity, distributed, scalable solution?

- Destination-aware vs destination unaware BC

- Ego-centric vs socio-centric computation of BC

CBC: the "destination-aware" counterpart to BC

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

Betweenness Centrality (*u*): portion of all pairs shortest paths of G that pass through node *u*

$$BC(u) = \sum_{s=1}^{|V|} \sum_{t=1}^{s-1} \frac{\sigma_{st}(u)}{\sigma_{st}}$$

Conditional **Betweenness Centrality (*u, t*)** : portion of all shortest paths of G from node *u* to *target t,* that pass through node *u*

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

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

The content placement problem

Deploy scalable and distributed mechanisms for publishing, placing, moving UG Service facilities / content within networking structures

Optimal content / service placement in a Graph ←→ k-median

Only distributed, scalable, solutions are relevant

- ➢ Use local information to migrate towards a better location
- ➢ Use locally available limited information to solve repeatedly small-scale k-median and repeat

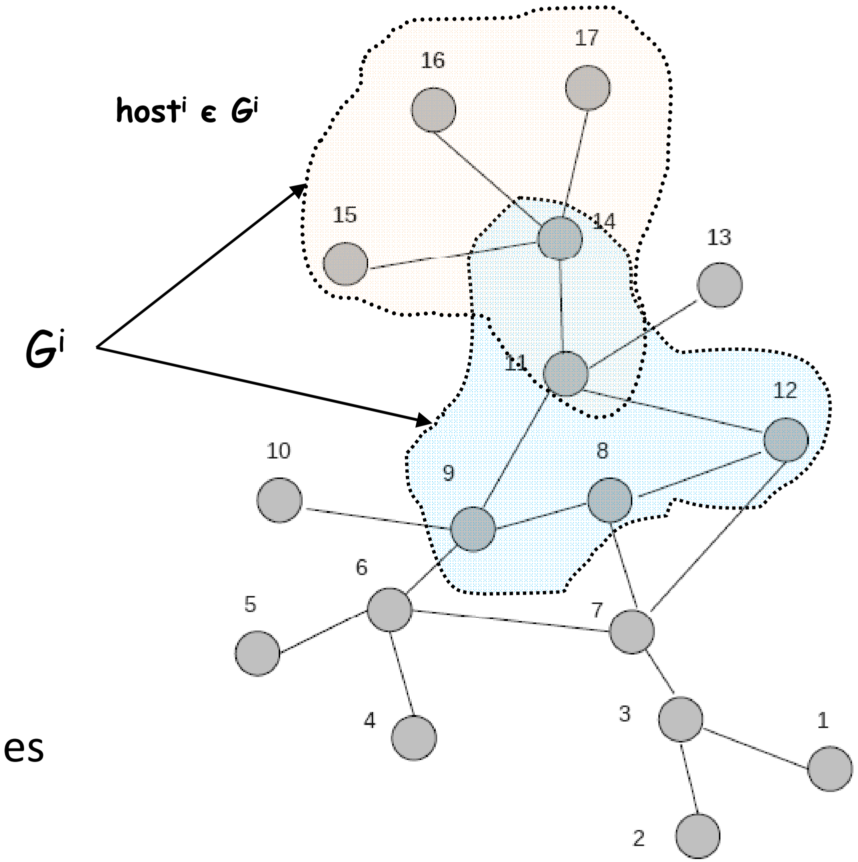
(*)

- K. Oikonomou, I. Stavrakakis, "Scalable Service Migration in Autonomic Network Environments," IEEE JSAC, Vol. 28, No. 1, Jan. 2010
- G. Smaragdakis, N. Laoutaris, K. Oikonomou, I. Stavrakakis, A. Bestavros, "Distributed Server Migration for Scalable Internet Service Deployment", to appear in IEEE/ACM T-Net. (2011) , also in INFOCOM2007

Centrality-based service migration

Consider set of nodes with highest CBC values

- ❑ Solve *iteratively* **small-scale** k-medians on subgraphs G^i **ϵ** G, around the current facility location of host i containing the <u>top nodes based on CBC values</u>



- ❑ Map the outside demand properly on nodes in subgraphs G^i I

P. Pandazopoulos, M. Karaliopoulos, I. Stavrakakis, "Centrality-driven scalable service migration", 23rd International Teletraffic Congress (ITC), Sept. 6-9, 2011, San Francisco, USA.

12

simulation results: ISP topologies / non-uniform load

Less than a dozen of nodes is enough!

Demand load : Zipf distribution (with skewness s)

Datasets correspond to different snapshots of 7 ISPs collected by `mrinfo` multicast tool *

$$\beta_{alg}(\alpha; G, \overline{w}) = E[\ \frac{C_{alg}(\alpha; G, \overline{w})}{C_{opt}(\alpha; G, \overline{w})}\ ]$$

$$\alpha_{\epsilon} = argmin\ \{\alpha | \beta_{alg}(\alpha) \leq\ (1+\epsilon)\}$$

Table 1.2 Results derived by the Generalized Neighborhood Service Migration Strategy

Size of physical topology	Min. Subgraph size for solutions within 2.5% of the optimal	
	$s=0$	$s=1$
76	4	4
100	5	5
180	5	4
184	4	4
216	4	4
339	7	6
378	5	5

***** J.-J. Pansiot, P. Mérindol, B. Donnet, and O. Bonaventure, "Extracting intra-domain topology from mrinfo probing," in Proc. Passive and Active Measurement Conference (PAM), April 2010.

13

Ego-centric vs socio-centric computation of BC

Very high **rank** correlation (Spearman coefficient) !!!

➔Ego- and socio- centric metrics identify same subsets

TABLE IV
CORRELATION STUDY BETWEEN BC-EGOBC AND CBC-EGOCBC ON INTRA-DOMAIN ISP TOPOLOGIES

	DataSet	ISP(AS number)	<CC>	Diameter	Size	<degree>	BC vs. ego-BC				CBC vs. ego-CBC	
							Spearman ρ		Pearson r_{Prs}		Spearman ρ	
							ego-net. **r=1**	ego-net. **r=2**	ego-net. **r=1**	ego-net. **r=2**	ego-net. **r=1**	95%
Tier 1	36	Global Crossing(3549)	0.546	10	76	3.71	0.9648	0.9806	0.6720	0.9197	0.9568	0.008
	35	-//-	0.479	9	100	3.78	0.9690	0.9853	0.7029	0.9255	0.9489	0.013
	33	NTTC-Gin(2914)	0.307	11	180	3.53	0.9209	0.9565	0.7479	0.8561	0.9554	0.003
	21	Sprint(1239)	0.298	12	216	3.07	0.9718	0.9812	0.7470	0.8557	0.9824	0.002
	13	Level-3(3356)	0.169	25	378	4.49	0.2708	0.9393	-0.0918	0.7982	0.7336	0.007
	12	-//-	0.149	28	436	4.98	0.2055	0.9381	-0.1217	0.7392	0.7035	0.005
	20	Sprint(1239)	0.287	16	528	3.13	0.9866	0.9928	0.5805	0.8488	0.9847	0.003
	9	-//-	0.251	13	741	3.29	0.9901	0.9930	0.7149	0.8622	0.9884	0.002
Transit	40	JanetUK(786)	0.132	14	336	2.69	0.9714	0.9825	0.8049	0.9180	0.9819	0.001
	45	Iunet(1267)	0.246	11	598	3.88	0.8506	0.9468	0.8887	0.9688	0.7825	0.033
	38	-//-	0.231	12	645	3.75	0.8790	0.9516	0.9094	0.9568	0.8062	0.022
	39	-//-	0.038	13	711	3.45	0.9470	0.9826	0.5354	0.9536	0.9370	0.016
	44	Telecom Italia(3269)	0.037	13	995	3.65	0.7950	0.9828	0.3362	0.8699	0.9902	0.001

P. Pandazopoulos, M. Karaliopoulos, I. Stavrakakis, “Egocentric assessment of node centrality in physical network topologies”, submitted to Globecom 2011

2- Betweeness Centrality (BC)

2.1 - Centrality-driven routing in opportunistic nets

(SimBetTS and BubbleRap use BC values of encounters for content forwarding)

How is performance of centrality-based routing affected by

- ❑ Adding or not, destination awareness to BC (BC vs CBC)
- ❑ Working with ego-centric vs socio-centric BC values
- ❑ Type of contact graph (unweighted vs. weighted) ? Not discussed here

P. Nikolopoulos, et.al. "How much *off-center are centrality metrics for* opportunistic routing?", CHANTS 2011 Workshop (in MobiCom), Sept 23, 2011, Las Vegas

Datasets

5 well-known iMote-based real traces available from the Haggle Project at CRAWDAD.

CHARACTERISTICS OF EMPLOYED DATASETS

Configuration	Intel	Cambridge	Infocom05	Content	Infocom06
Device type	iMote	iMote	iMote	iMote	iMote
Network type	B/T	B/T	B/T	B/T	B/T
Duration (days)	6	6	4	24	4
Scan time (sec)	5-10	5-10	5-10	5-10	5-10
Granularity (sec)	120	120	120	120-600	120
Mobile Devices	8	12	41	36	78
Stationary Dev.	1	0	0	18	20
External Dev.	119	211	233	11368	4421
Average internal contacts/pair/day	9.09	12.09	8.60	0.66	9.03
# of Contacts	2766	6732	28216	41330	227657

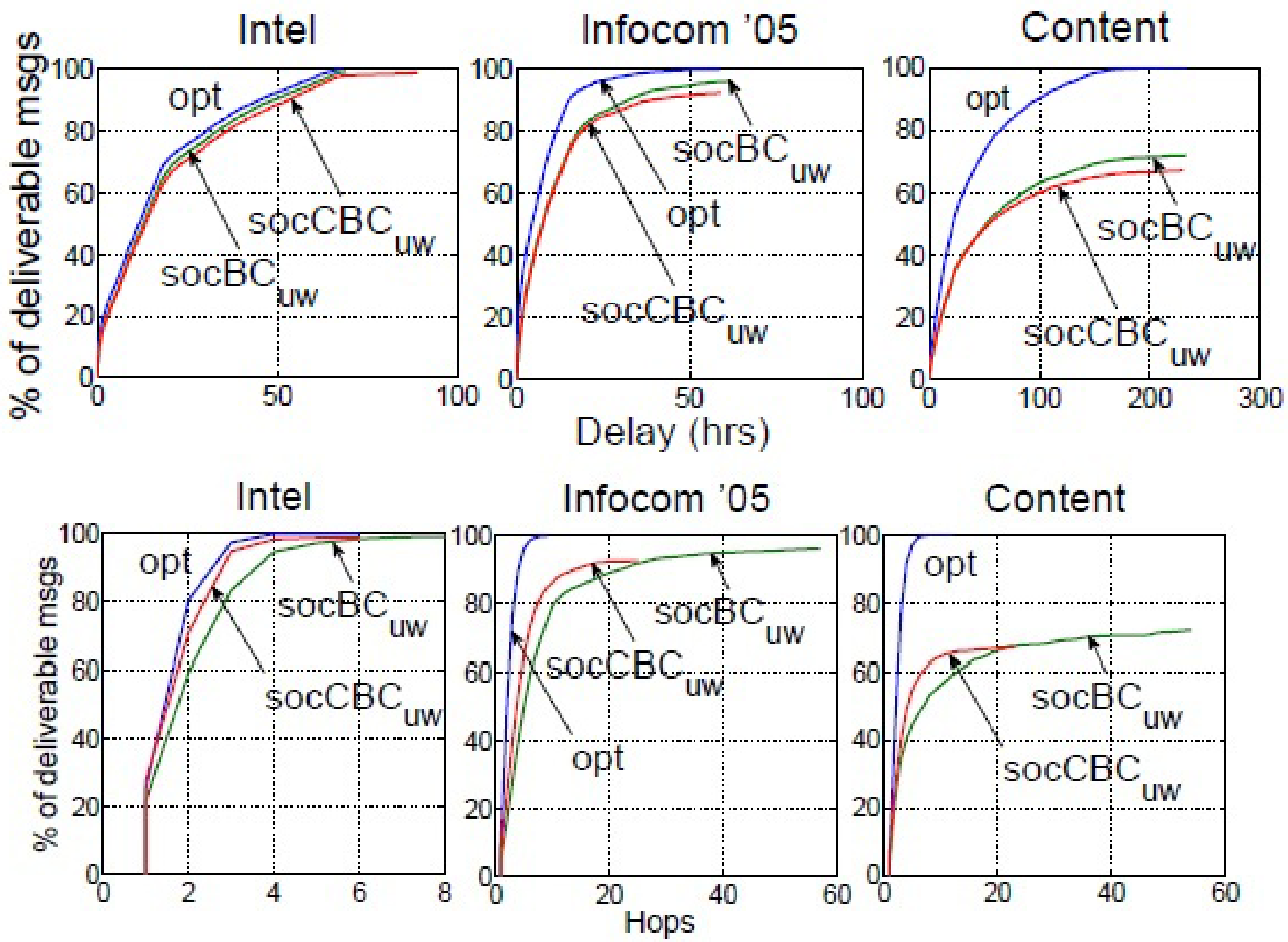
BC vs CBC

- opt ➔ optimal routing through knowledge of contact sequences.

- BC/CBC ➔ up to 30% of messages never reach their destination
  ➔ about 5 times more hops and 1 day of additional delay

BC outperforms CBC in delay (due to zero CBC values when destination in an unconnected cluster)

CBC outperforms BC in hops (up to 50% shorter paths, due to selecting more proper nodes to forward to)



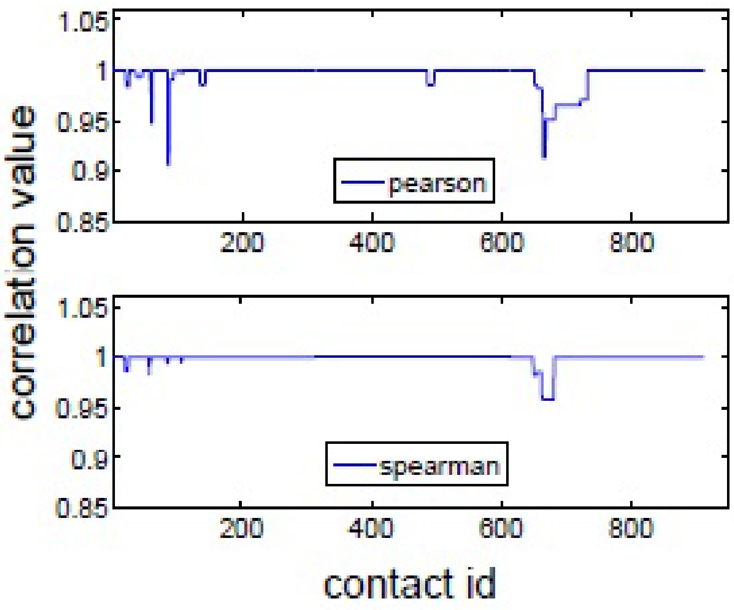
socio- vs ego-metrics

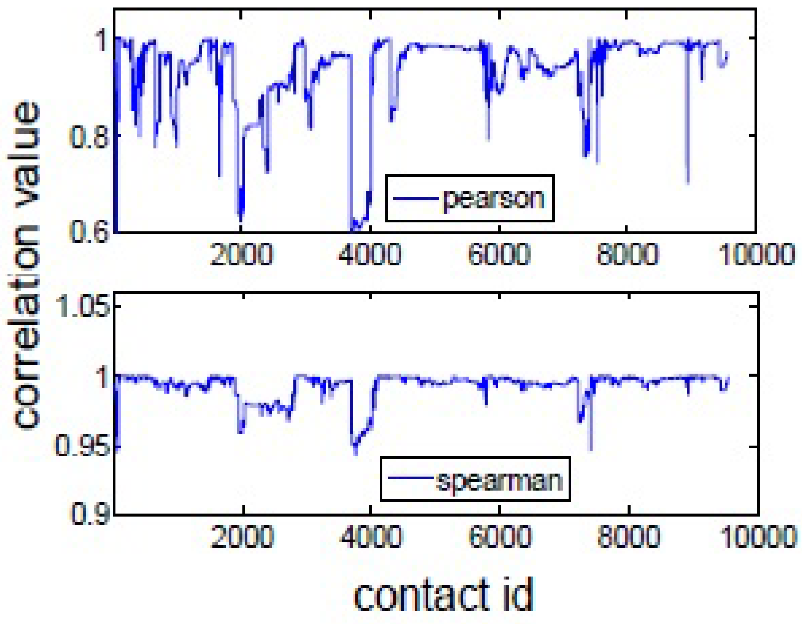
Table 2: Probability of delivery

DataSet	Probability of delivery (6h window)			
	$egoBC_{uw}$	$socBC_{uw}$	$egoBC_{w}$	$socBC_{w}$
Intel	99.26	99.26	99.24	99.24
Cambridge	82.68	84.50	82.14	77.24
Infocom'05	95.21	96.25	88.06	86.40
Content	65.55	71.84	69.18	69.20
Infocom'06	80.93	85.08	89.33	89.73

socio- vs ego-metrics

strong positive correlation of socio- and ego – metrics
(Intel / Content data)





Conclusions

Focused on exploring the impact on two key social metrics on content distribution

- Interest Similarity
- Centrality

Interest similarity metrics

- Highly similar groups can yield high gains in content replication.
- Interest similarity –based forwarding improves performance
- Worth assessing interest similarity in groups – framework for doing that

Destination-aware BC :

- Very effective in content placement (BC is totally ineffective)
- Decreases hop count in opp nets (energy) substantially. Can increase delay

Ego-centric centrality variants (BC/CBC)

- Highly rank correlated → no performance degradation in content placement / centrality-driven content forwarding.
- Easier to compute