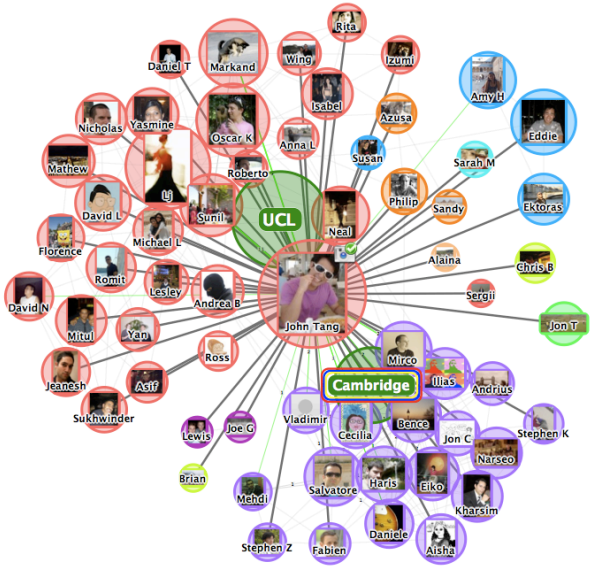
UNIVERSITY OF CAMBRIDGE
Computer Laboratory



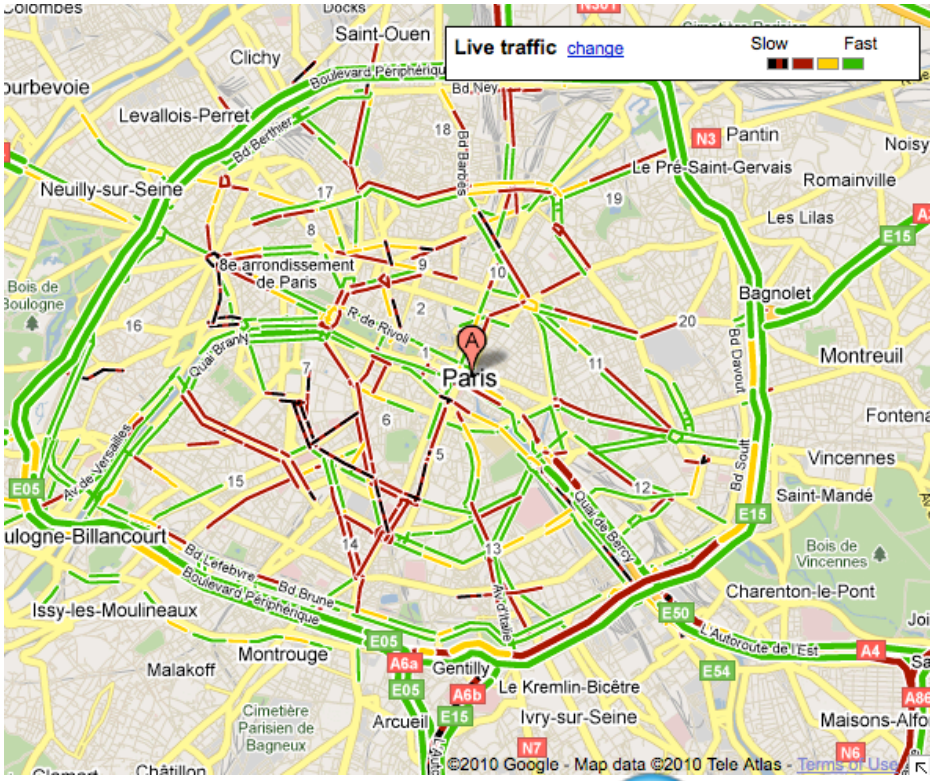
Studying mobile and social networks through time-aware network analysis

John Tang
Computer Lab, University Of Cambridge

With with: Vito Latora, Cecilia Mascolo, Mirco Musolesi and Vincenzo Nicosia

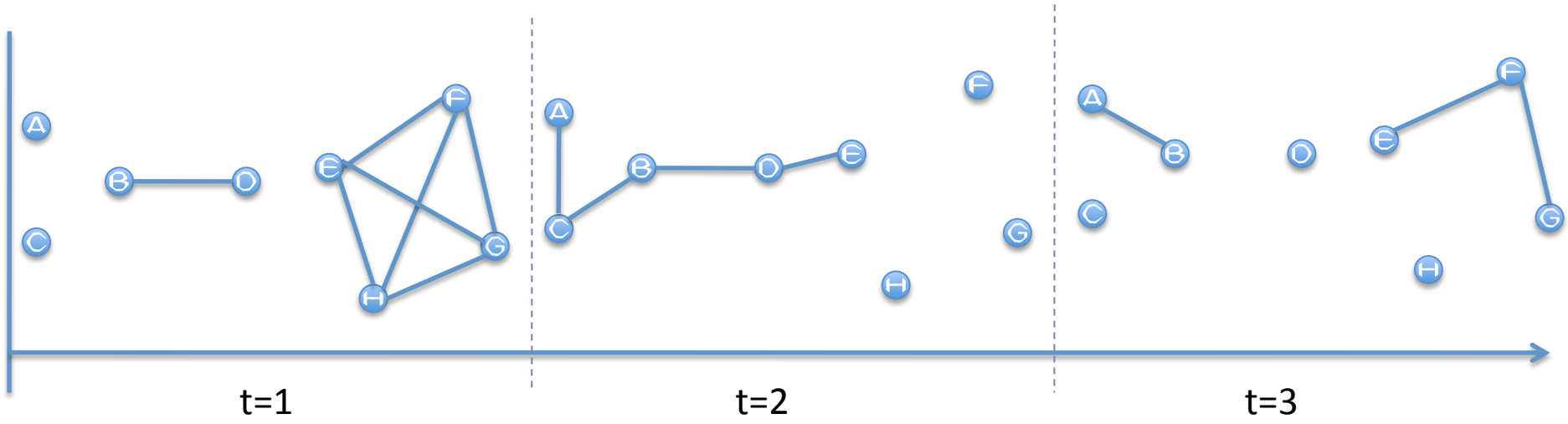
The 9th Mathematics of Networks, 19th June 2010, University of St. Andrews

Some Real Networks

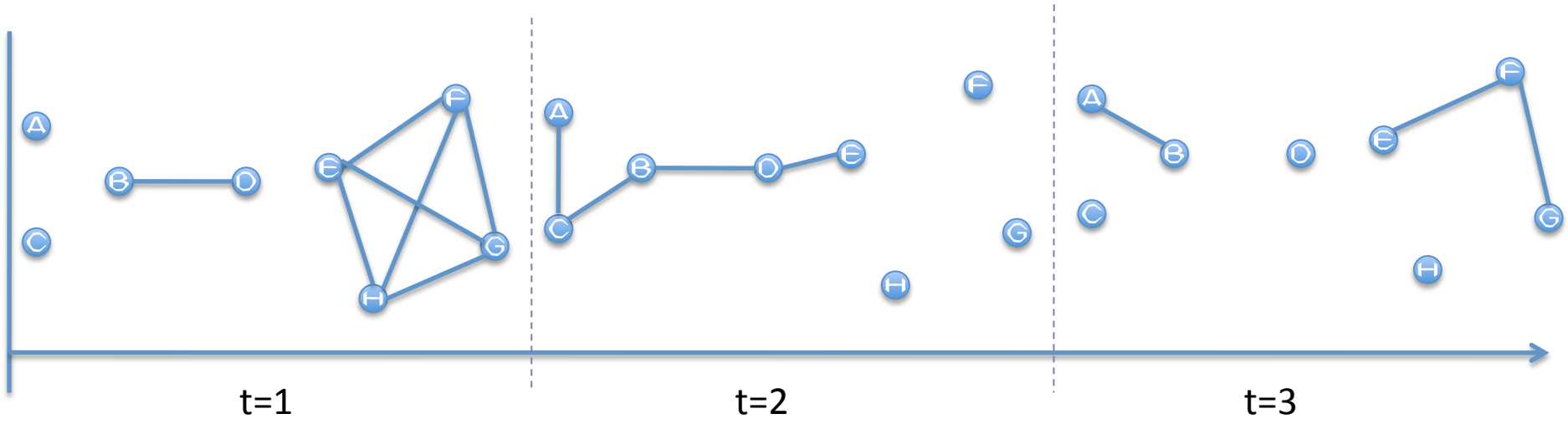


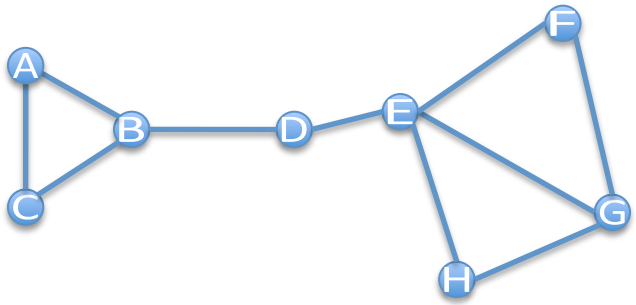
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Temporal Graph

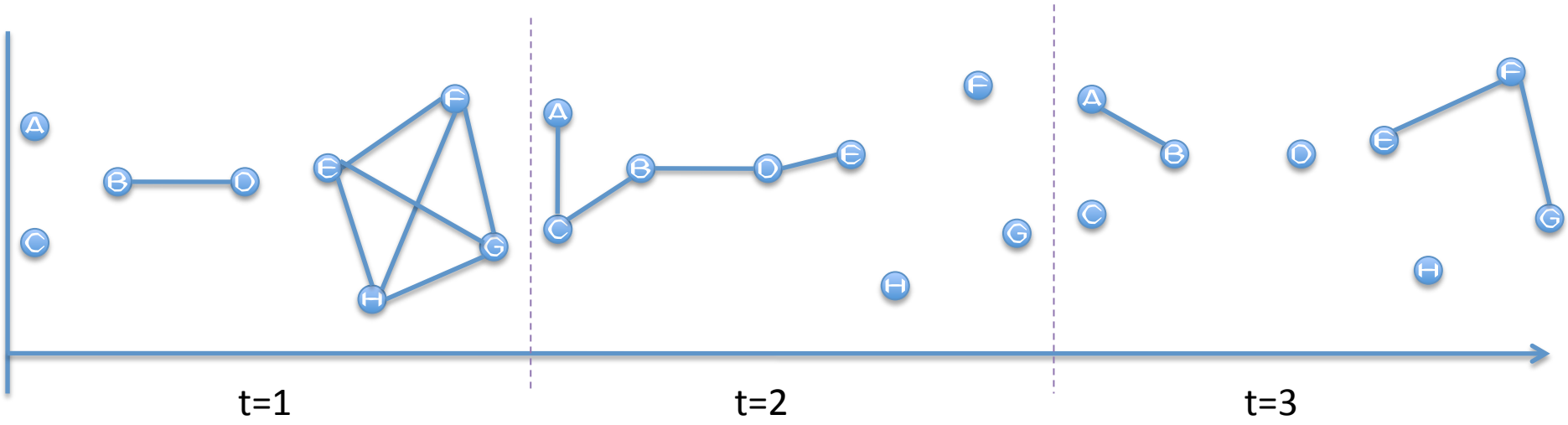


Temporal Graph

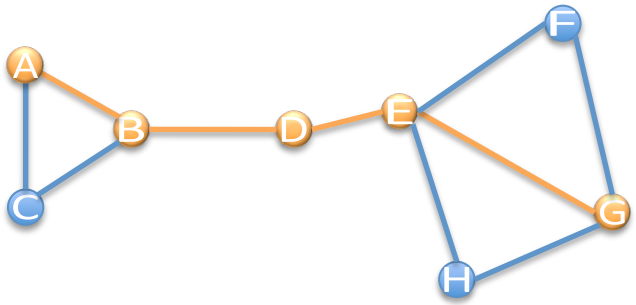




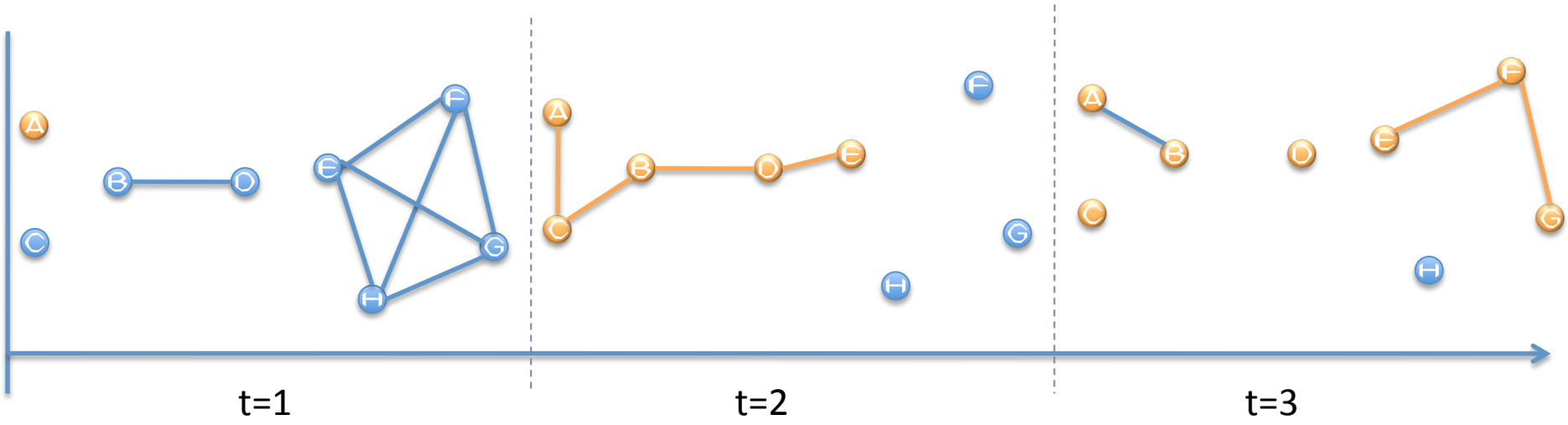
Temporal Graph



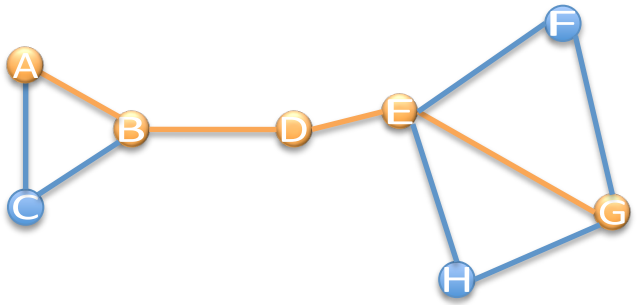
- Static
  - Shortest path (A,G) = [A,B,D,E,G]
  - Shortest path length (A,G) = 4 hops



Temporal Graph



- Static
  - Shortest path (A,G) = [A,B,D,E,G]
  - Shortest path length (A,G) = 4 hops
- Temporal
  - Shortest path (A,G) = [A,C,B,D,E,F,G]
  - Shortest path length (A,G) = 6 hops
  - Time=3 seconds



Temporal Metrics

- d_{ij} Shortest Temporal Path Length
- d^*_{ij} Shortest Path with temporal constraints
- $E_{ij} = \frac{1}{d_{ij}}$ Temporal Efficiency

Temporal Metrics

- Average Temporal $L = \frac{1}{N(N-1)} \sum_{ij} d_{ij}$
- Average Temporal $L^* = \frac{1}{N(N-1)} \sum_{ij} d^*_{ij}$
- Average Efficiency $E_{glob} = \frac{1}{N(N-1)} \sum_{ij} E_{ij}$

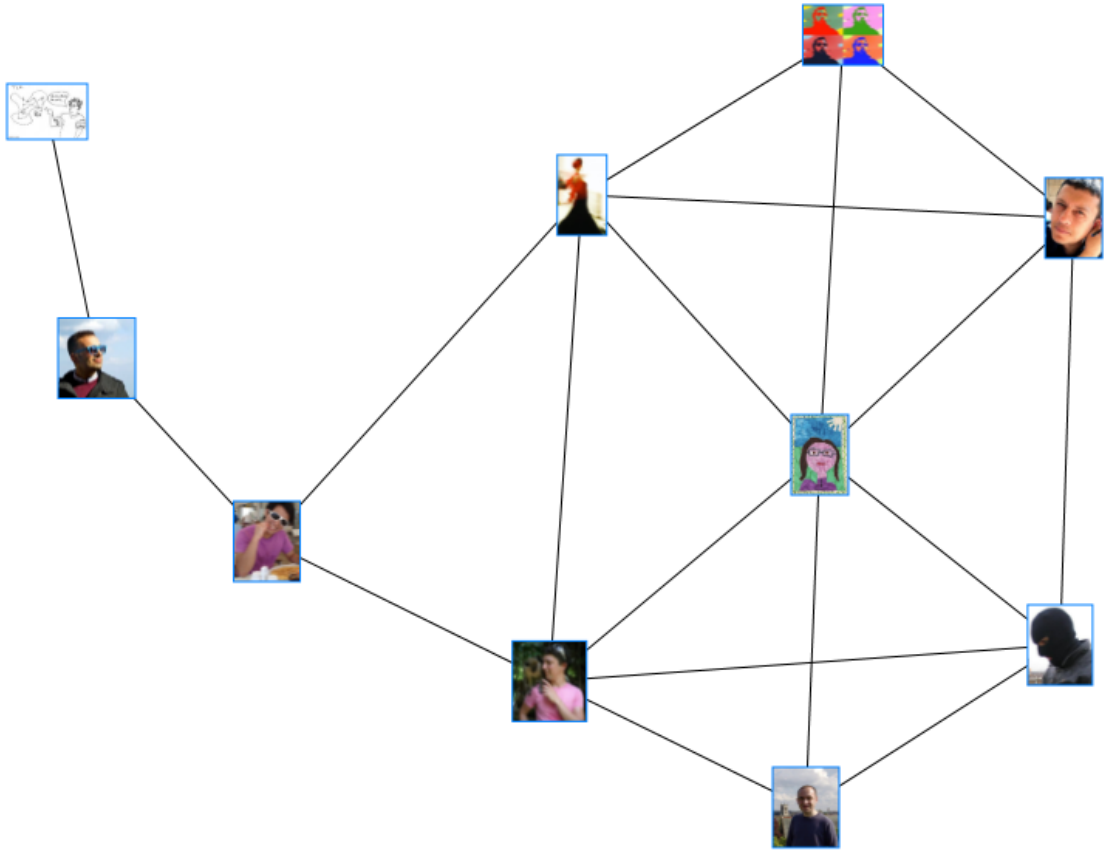
Does it really matter?

- Infocom 2005 conference environment
- Bluetooth colocation scans
- 5 Minute Windows
- Measure 24 hours starting 12am

					Static		Temporal		
Day	**N**	**<k>**	**Activity**	**Contacts**	**L**	**Eglob**	**L***	**L**	**Eglob**
1	37	25.73	6pm-12pm	3668	1.291	0.856	4.090	19h 39m	0.003
2	39	28.31	12am-12pm	8357	1.269	0.870	4.556	9h 6m	0.024
3	38	22.32	12am-12pm	4217	1.420	0.798	4.003	10h 32m	0.018
4	39	21.44	12am-5pm	3024	1.444	0.781	4.705	9h 55m	0.013

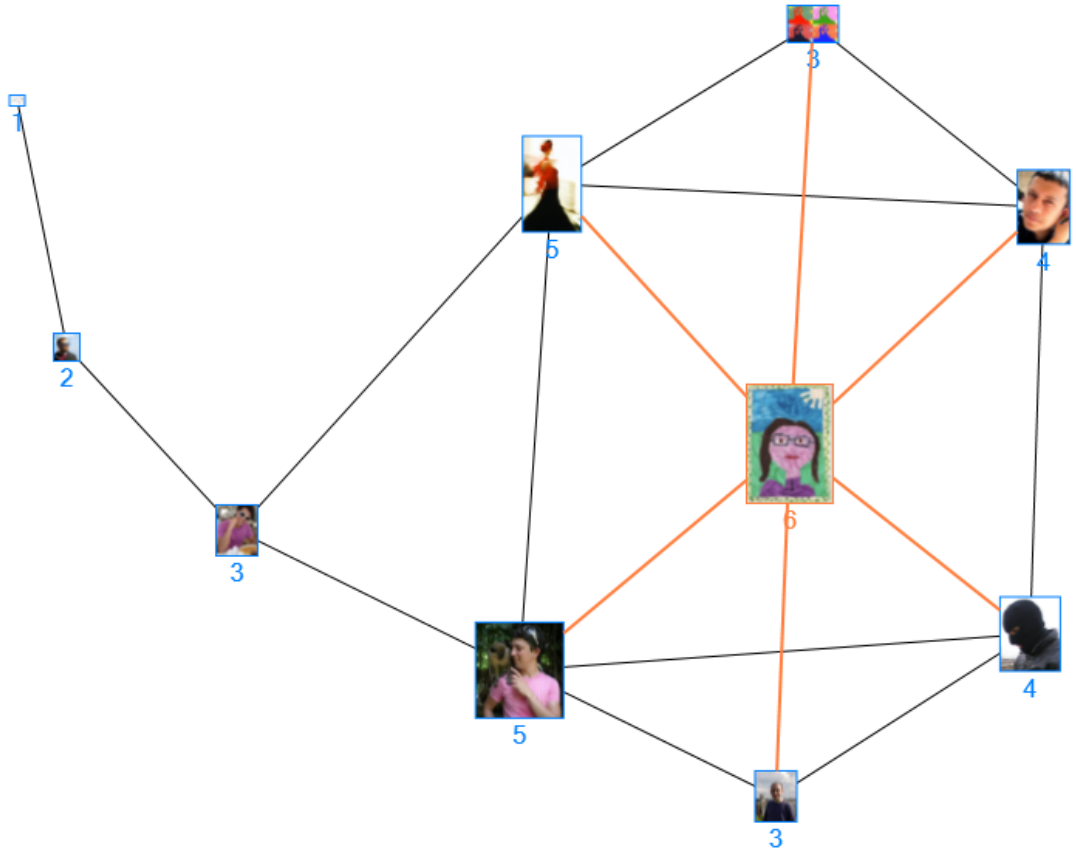
Important Nodes

- Most number of friends
- Quickly spread information to many people
- Mediates between the most information flows



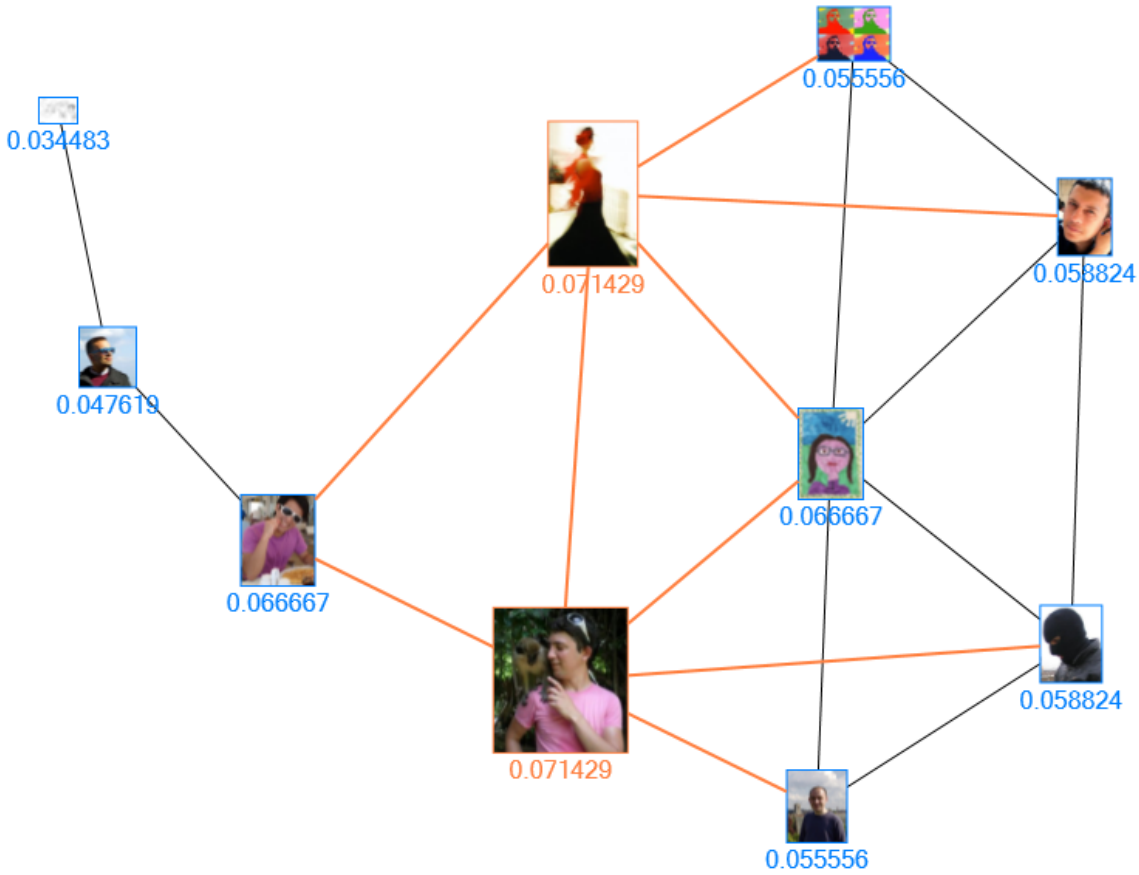
Degree

- $C_i^{deg} = \text{number of links to i}$
- Popular nodes



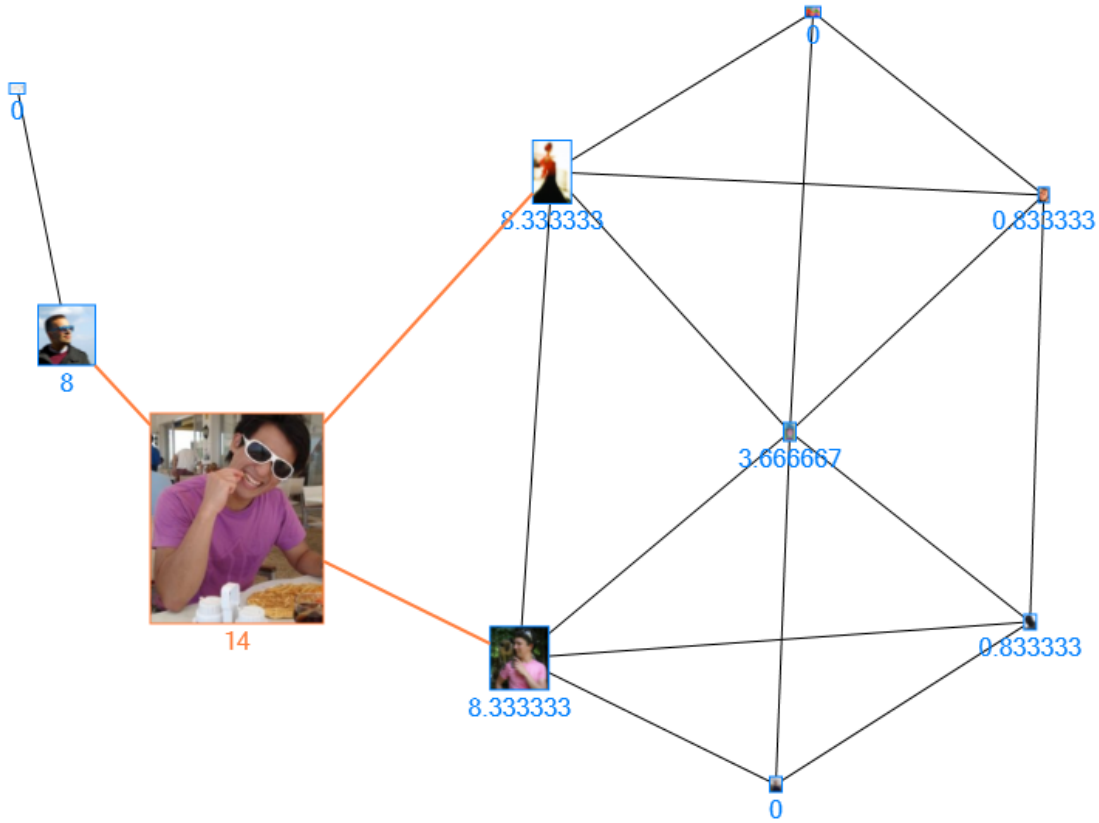
Static Closeness Centrality

- $C_i = \sum_{i \neq j} d_{ij}$
- Average shortest path length to all other nodes



Static Betweenness Centrality

- $C_i^{bet} = \sum_{i \neq s \neq t} \frac{\delta_s t(i)}{\delta_{st}}$ where δ_{st} is # shortest paths from s to t
  $\delta_{st}(i)$ is # shortest paths passing through i
- Fraction of shortest paths which pass through node i



Temporal Centrality Metrics

- Static Closeness and Betweenness based on *static shortest paths*
- Reformalise *closeness* and *betweenness* with temporal paths:
  - Duration
  - Time Order
  - Frequency

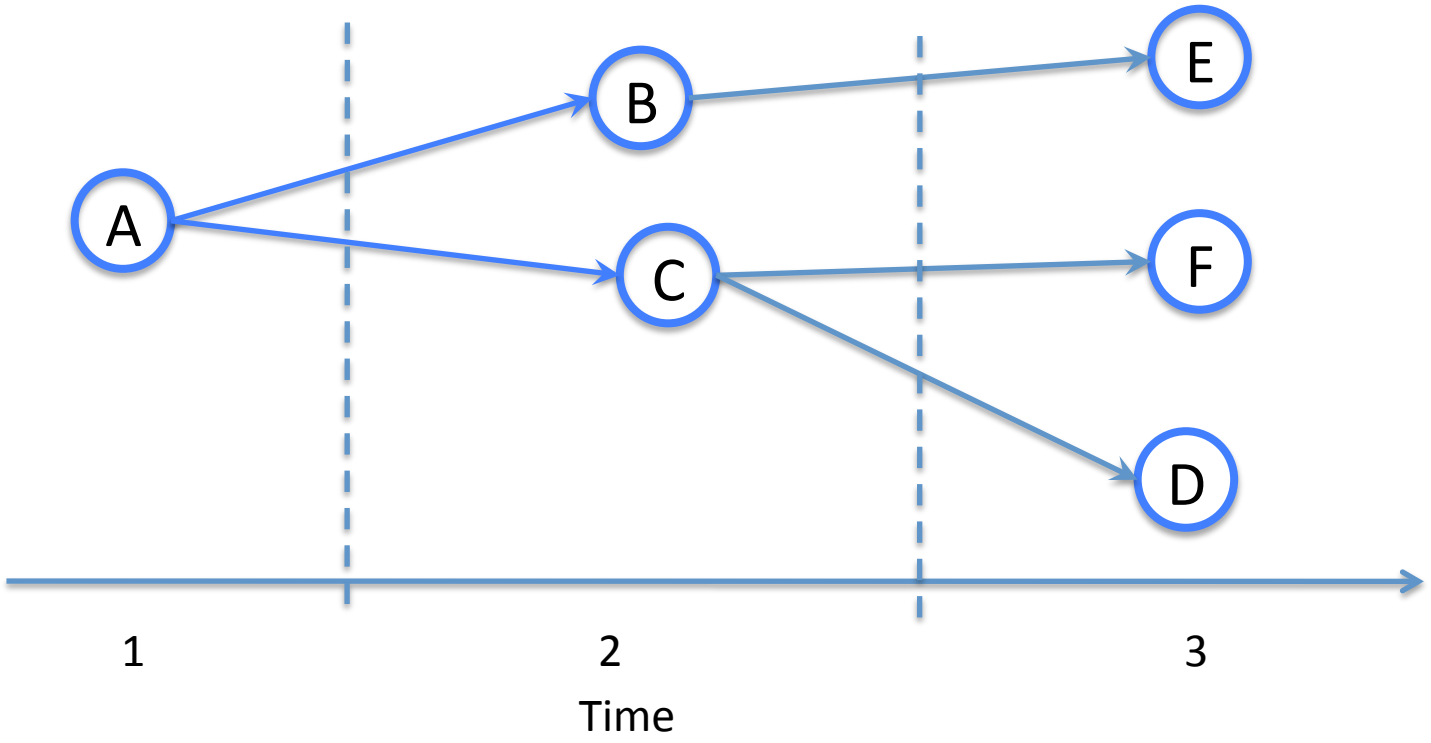
Temporal Closeness

Average over shortest *temporal* paths to all other nodes:

$$C_i = \frac{1}{W(N-1)} \sum_{j \neq i \in V} d_{i,j}$$

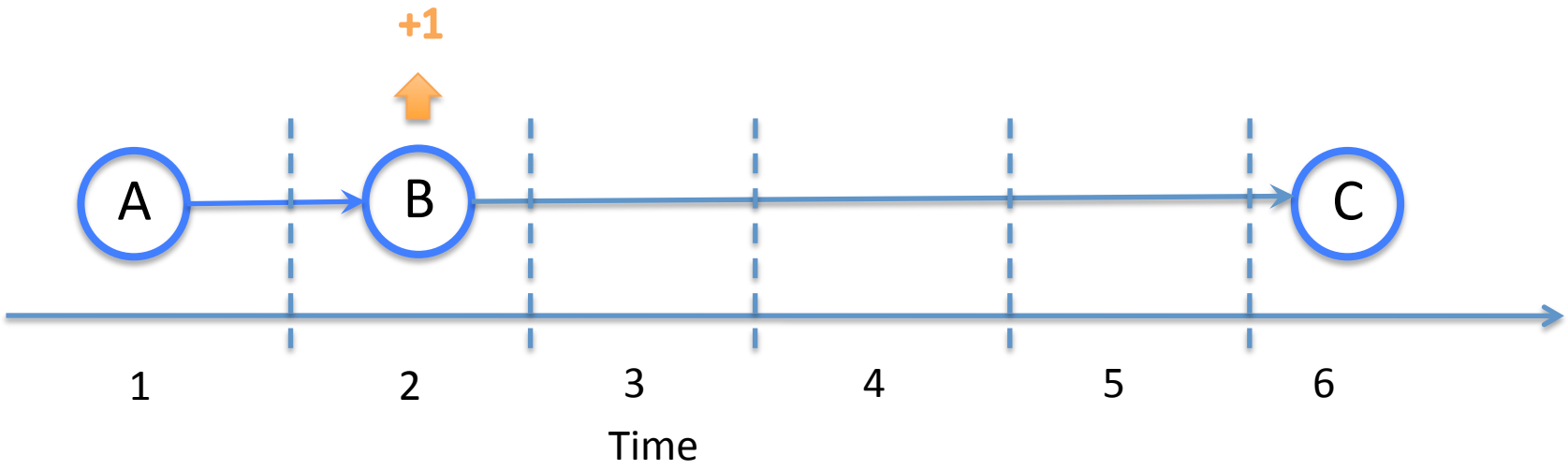
Temporal Closeness

$$C_A = \frac{(2+2)+(3+3+3)}{(3*(6-1))} = 0.867$$



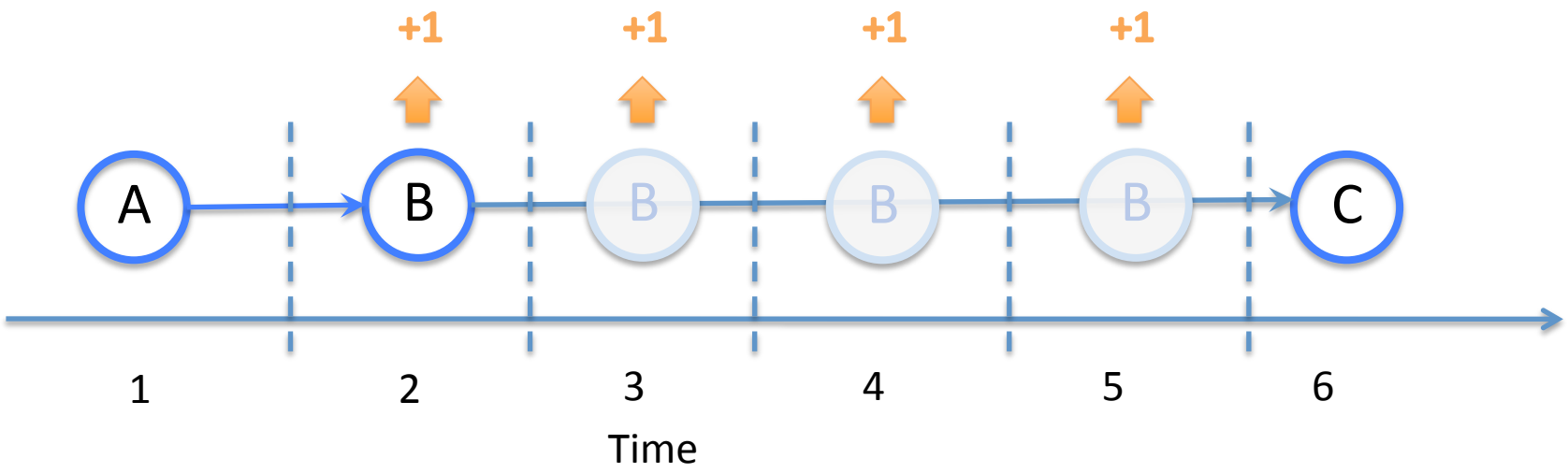
Temporal Betweenness

- Using temporal path length



Temporal Betweenness

- Take into account **duration**



Temporal Betweenness

$$C_i^B(t) = \frac{1}{(N-1)(N-2)} \sum_{\substack{j \in V \\ j \neq i}} \sum_{\substack{k \in V \\ k \neq i \\ k \neq j}} \frac{U(i,t,j,k)}{|S_{jk}^h|}$$

Where:

- $U(i,t,j,k)$ returns number of shortest paths from j to k, which node i is holding a message at time window t
- $|S_{jk}^h|$ number of shortest temporal paths between j and k

Temporal Betweenness

Sum over all time windows for each node:

$$C_i^B = \frac{1}{W} \sum_{t=1}^{W} C_i^B((t \times w) + t_{min})$$

Evaluating Centrality

- Corporate Email Dataset
- Two perspectives:
  - Semantic: roles of each node
  - Dynamic Processes: simulate communication
    - Information Dissemination
    - Information Mediation

Evaluating Centrality

- Corporate Email Dataset
- Two perspectives:
  - Semantic: roles of each node
  - Dynamic Processes: simulate communication
    - Information Dissemination
    - Information Mediation

Enron in the News

Scandals


ENRON
BARGE DEAL
$12M
MERRILL LYNCH
LJM
BARGES
$7.525M
LIFE
NEWS
CELEBRITY
TRAVEL
ANIMALS
SPORTS
Return to Search Results
Civil Rights: Women in the Struggle
Taliban: Oppression and Resistance
Disasters at Sea
Troops Honor Fallen Comrades
EMAIL
RSS
RATE
PRINT
LINK TO
LICENSE
DIGG IT
TWEET
SHARE
California Experiences Second Week of Energy Crisis
384758 04: FILE PHOTO: Steam rises from the Etiwanda Generating Plant behind power lines near Rancho Cucamonga, CA, January 24, 2001, as the statewide energy crisis continues. The power plant is owned and operated by Edison O&M Service Division. Blackout warnings were issued for the first time by the California Independent System Operator June 17, reporting that blackouts could occure Monday and Tuesday afternoons between noon and eight in the evening.
Photo: David McNew/Getty Images
Jun 18, 2001

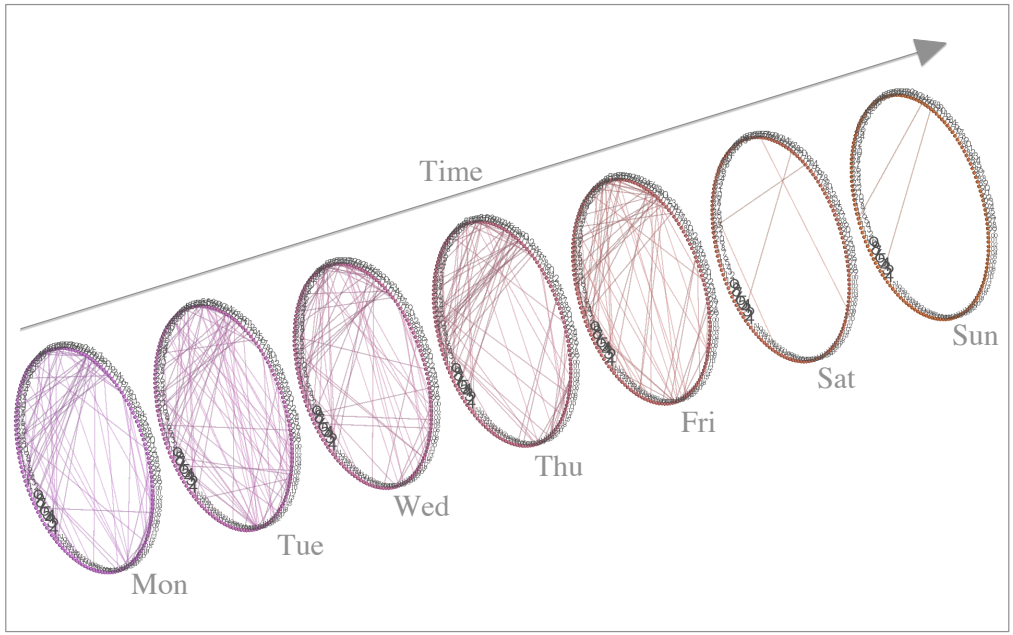
Public Investigation

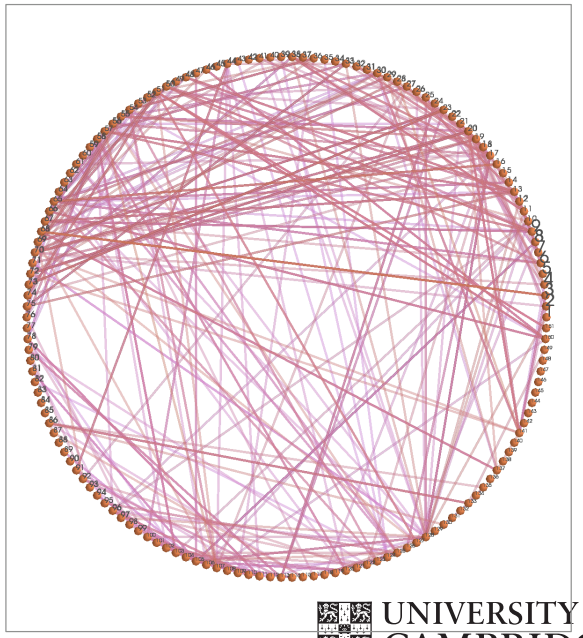
- Telephone logs
- Documents
- Financials
- Emails

  - 151 user mailboxes
  - May 1999 to Jun 2002
  - 250,000 emails
  - NOT anonymised

Email exchanges to Temporal Graph

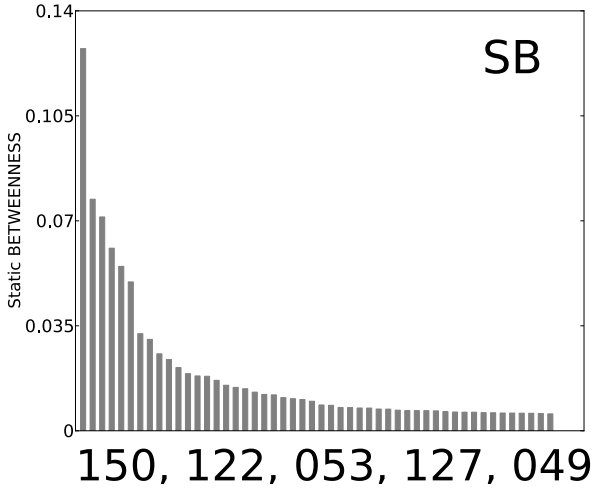
- Core 151 users
- Window size= 1 business day
- 1137 days

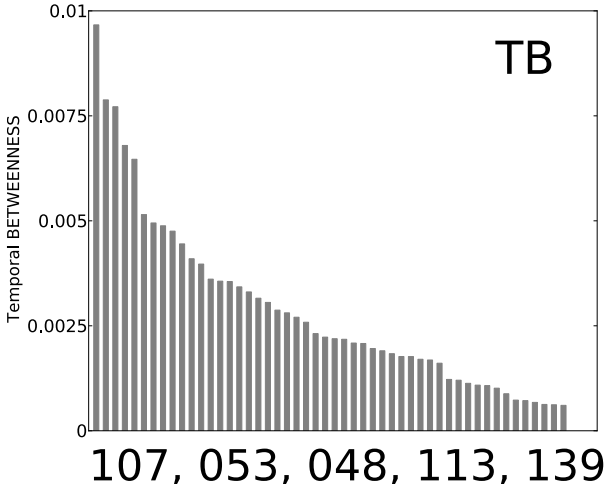


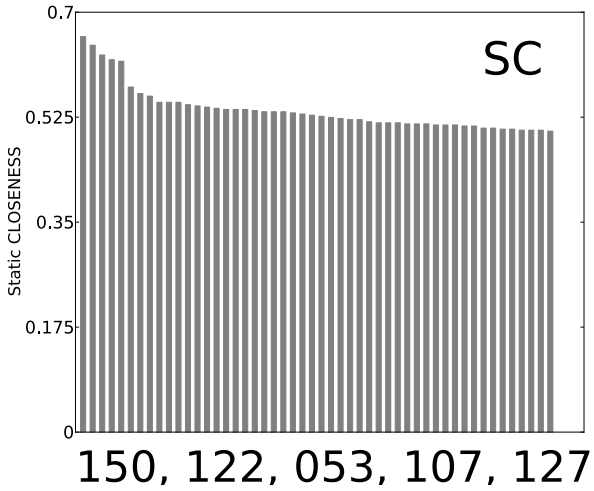


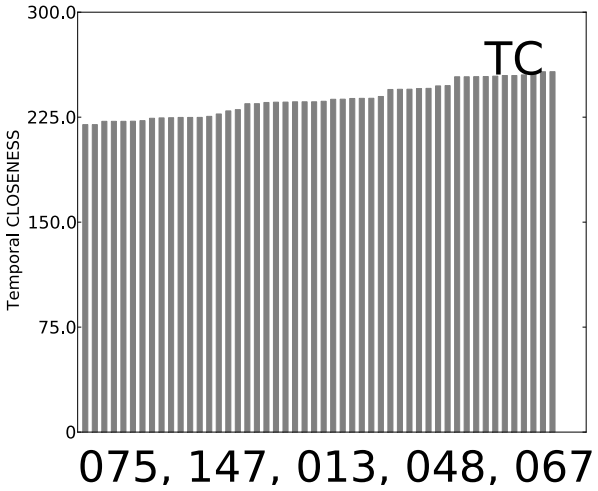
UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Centrality Rankings









UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Digging deeper

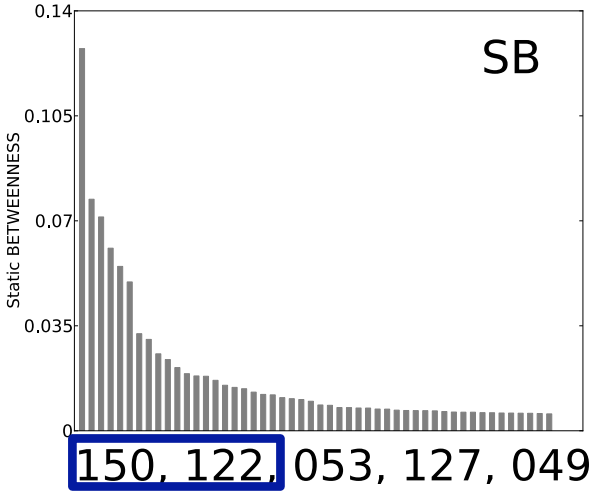
- Static centrality strongly correlated with degree

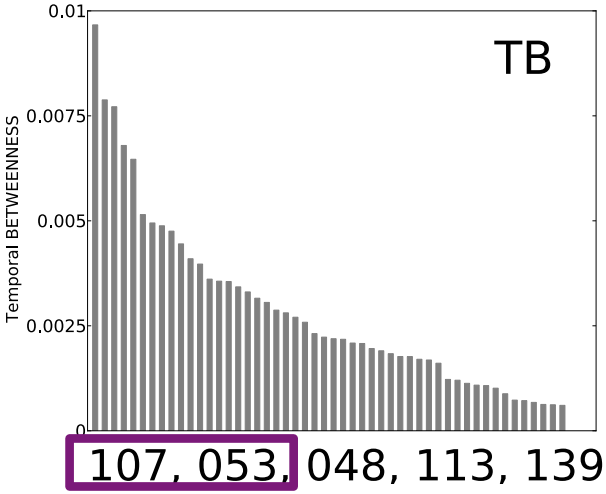
	SB	SC	SD	TB	TC	TD
SB	1.00	0.57	0.69	0.41	0.24	0.43
SC	-	1.00	0.70	0.36	0.22	0.31
SD	-	-	1.00	0.39	0.28	0.48
TB	-	-	-	1.00	0.43	0.34
TC	-	-	-	-	1.00	0.40
TD	-	-	-	-	-	1.00

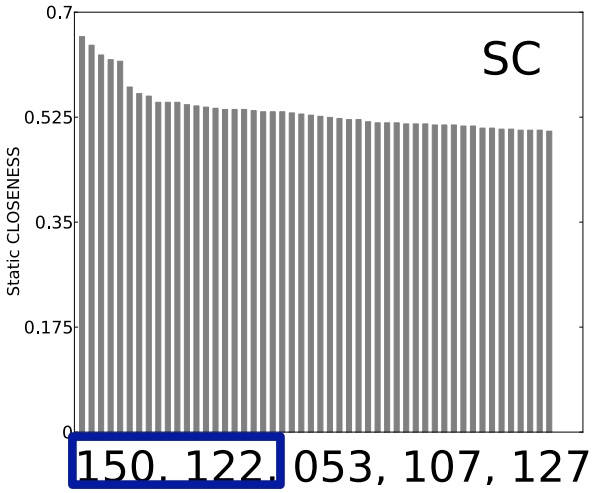
Evaluating Centrality

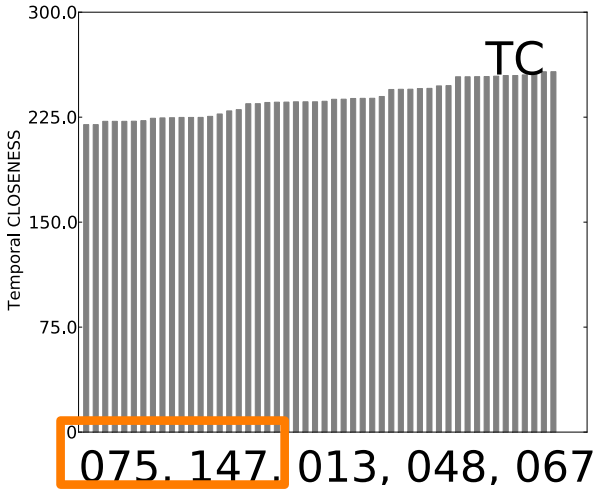
- Corporate Email Dataset
- Two perspectives:
  - Semantic: roles of each node
  - Dynamic Processes: simulate communication
    - Information Dissemination
    - Information Mediation

Semantics







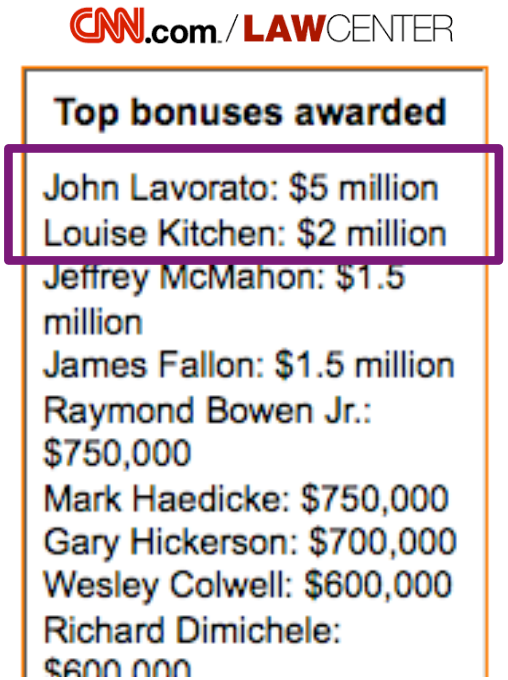


ID	Role
009	(Unknown)
013	Legal
017	Manager
048	Executive
053	Trader
054	President
067	Vice President
073	Trader
075	Director of Trading
107	Trader
122	Managing Director
127	Manager
139	Director
147	Trader
150	Secretary

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Semantics

ID	Name	Role
9	Stephanie Panus	(Unknown)
13	Marie Heard	Legal
17	Mike Grigsby	Manager
48	Tana Jones	Executive
53	John Lavorato	Trader
54	Greg Whalley	President
67	Sara Shackleton	Vice President
73	Jeff Dasovich	Trader
75	Gerald Nemec	Director of Trading
107	Louise Kitchen	Trader
122	Sally Beck	Managing Director
127	Kenneth Lay	Manager
139	Mary Hain	Director
147	Carol Clair	Trader
150	Liz Taylor	Secretary

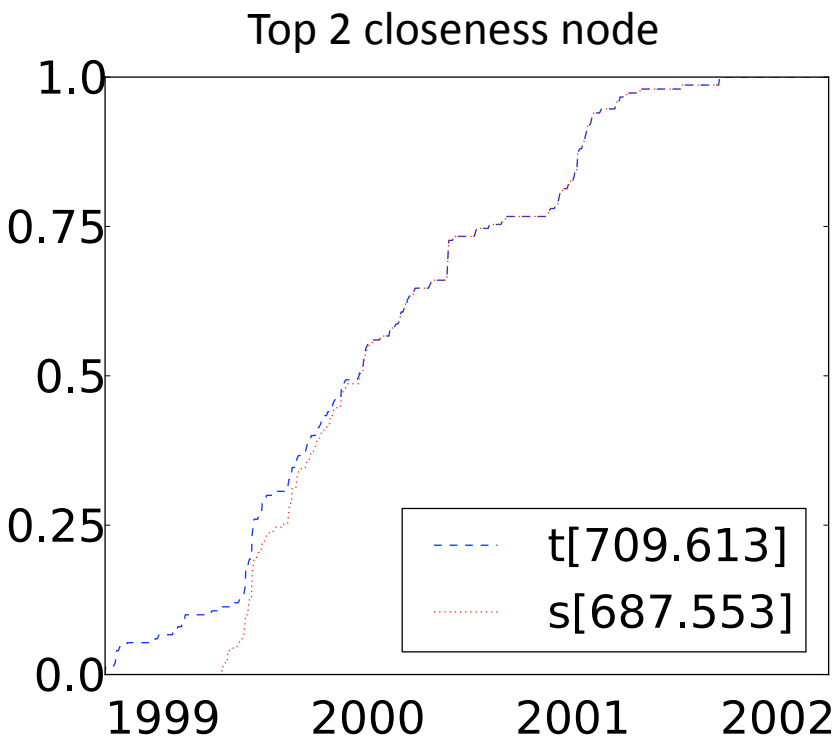


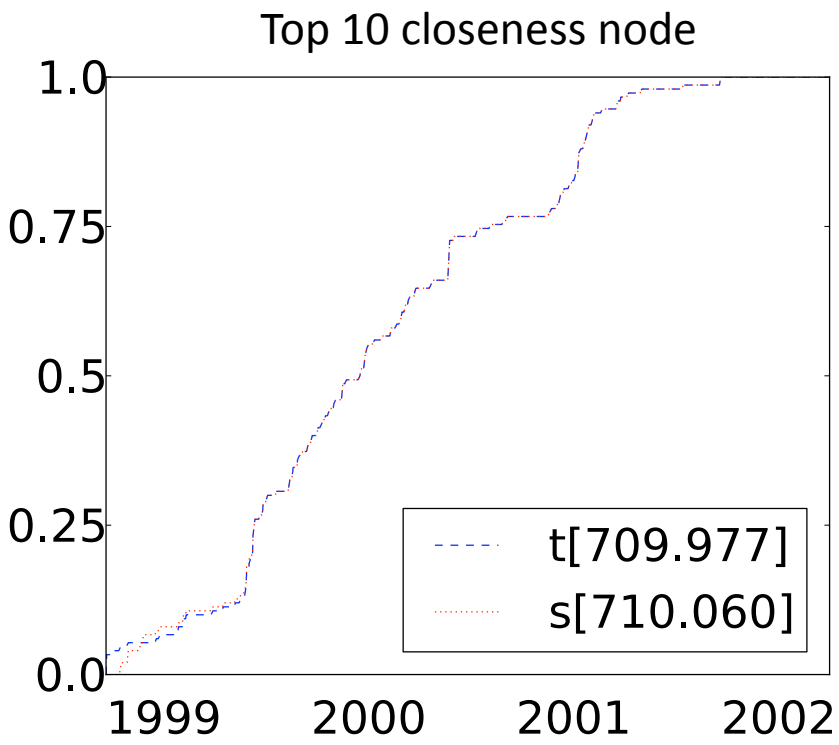
- Big bonuses linked with information mediators

Evaluating Centrality

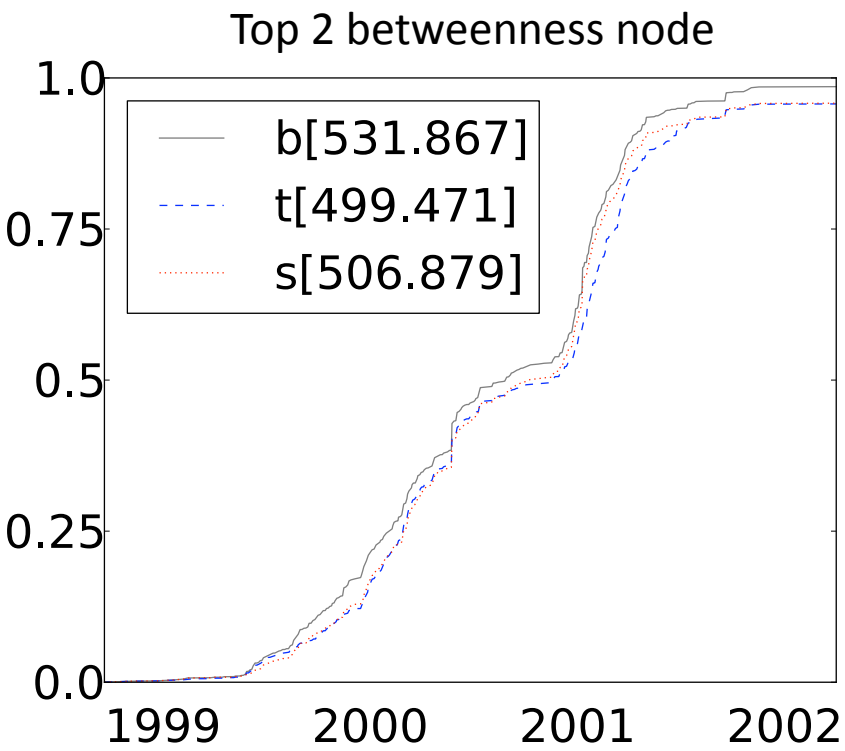
- Corporate Email Dataset
- Two perspectives:
  - Semantic: roles of each node
  - Dynamic Processes: simulate communication
    - Information Dissemination
    - Information Mediation

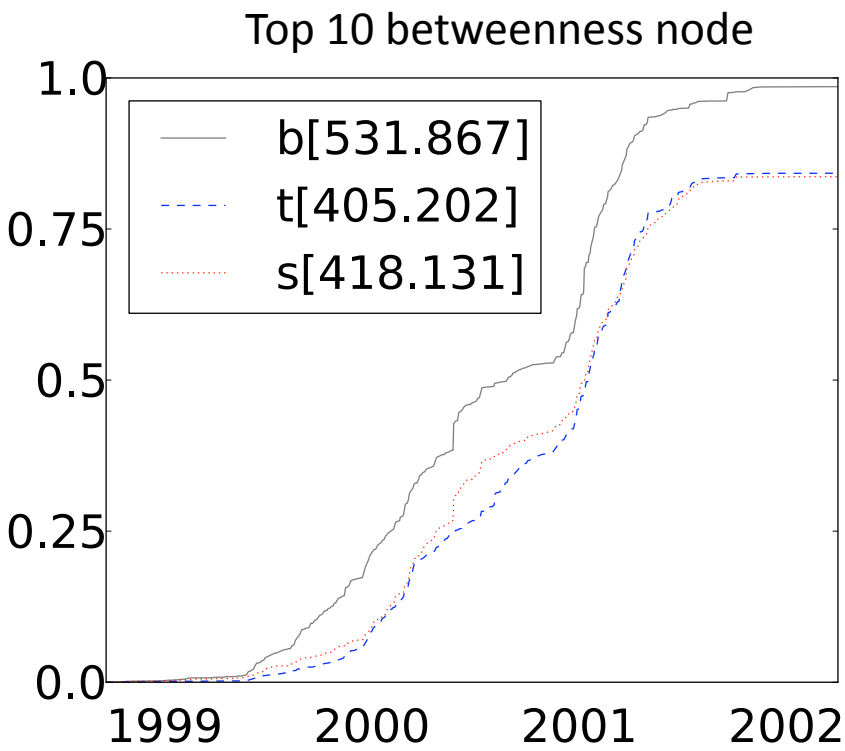
Information Dissemination



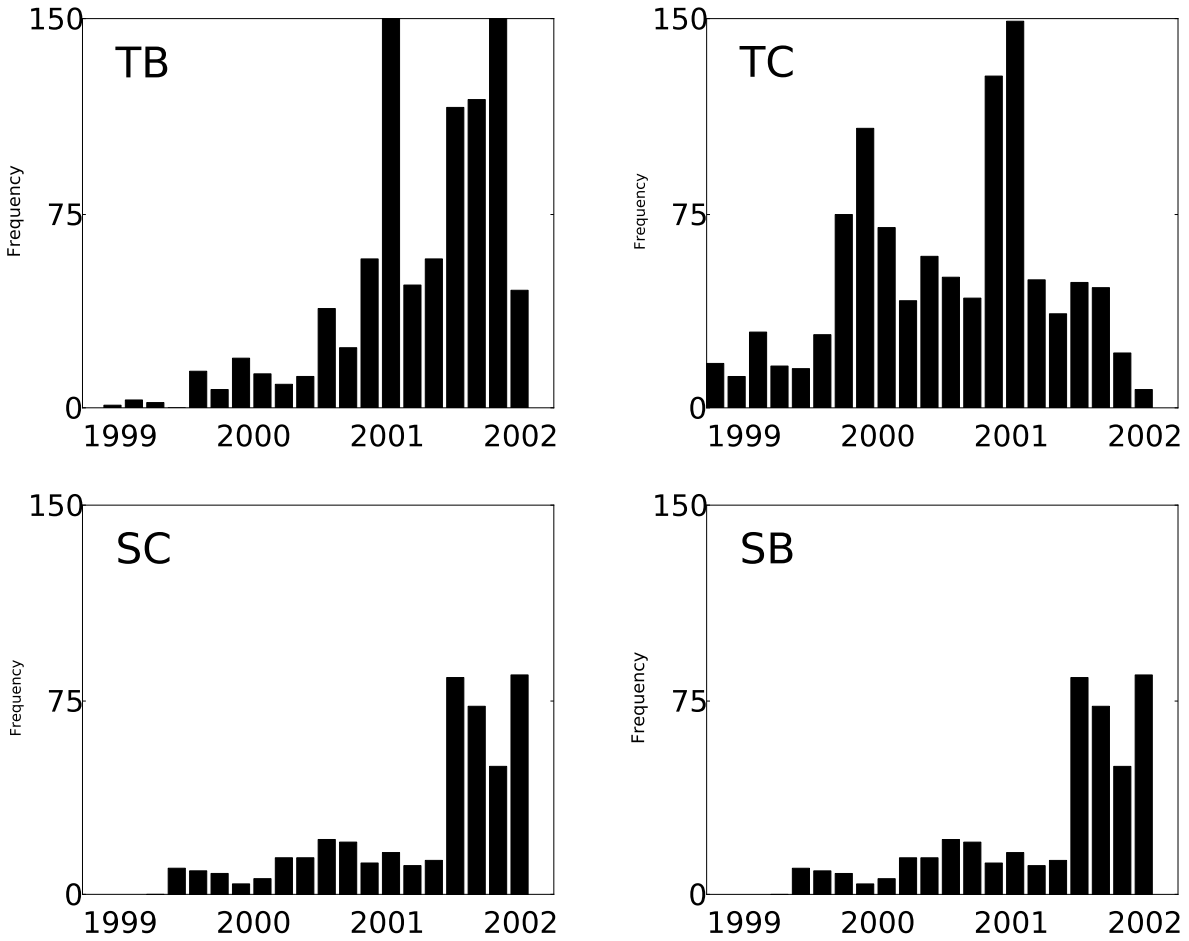


Information Mediation





Understanding Temporal Dynamics



UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Summary of Talk

- Temporal Graphs & Distance Metrics
  - Static shortest paths overestimate available hops and hence underestimate shortest path length
- Temporal Centrality Metrics:
  - More consistent node role
  - Better for dynamic processes
- Future Work
  - Inference using Temporal Model
    - Missing nodes/links, Future nodes/links, De-anonymise people
  - More traces with semantic information
  - Malware propogation
    - Best nodes for patching
  - Spectral Analysis

Questions?

John Tang

email jkt27@cam.ac.uk

homepage www.cl.cam.ac.uk/~jkt27

twitter @johnkittang

project http://www.cl.cam.ac.uk/research/srg/netos/**spatialtemporalnetworks**

Further Reading

Small World Behavior in Time-Varying Graphs, J. Tang, S. Scellato, M. Musolesi, C. Mascolo, V. Latora, Physical Review E, Vol. 81 (5), 055101, May 2010.

Analysing Information Flows and Key Mediators through Temporal Centrality Metrics, J. Tang, M. Musolesi, C. Mascolo, V. Latora, V. Nicosia, In Proceedings of the 3rd ACM EuroSys Workshop on Social Networks Systems (SNS10). Apr 2010.

Characterising Temporal Distance and Reachability in Mobile and Online Social Networks, J. Tang, M. Musolesi, C. Mascolo, V. Latora, ACM SIGCOMM Computer Communication Review (CCR). Vol. 40 (1), pp. 118-124. Jan 2010.

Temporal Distance Metrics for Social Network Analysis, J. Tang, M. Musolesi, C. Mascolo, V. Latora, In Proceedings of the 2nd ACM SIGCOMM Workshop on Online Social Networks (WOSN09). Aug 2009

UNIVERSITY OF CAMBRIDGE
Computer Laboratory