Metrics for temporal graphs

V. Nicosia[1,2]
J.K. Tang [1] C. Mascolo [1]
V. Latora [2,3,4]
Liam McNamara [5]
Mirco Musolesi[6]

[1]Computer Laboratory,University of Cambridge, UK
[2]Laboratorio sui Sistemi Complessi, Scuola Superiore di Catania, Italy
[3]School of Mathematical Sciences, Queen Mary College, University of London, UK
[4]Dipartimento di Fisica, Università di Catania, Italy
[5]IT Department, Communication Research Group, Uppsala Universitet, Sweden
[5]School of Computer Science, University of Birmingham, UK

Sep. 19 2012 – Cambridge

Overview

Classical network data

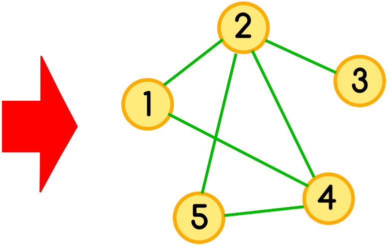
1^{st} Unit	2^{nd} Unit	**(Weight)**
1	2	**3**
1	4	**1**
2	3	**5**
2	4	**2**
2	5	**7**
4	5	**3**

Classical network data

1^{st} Unit	2^{nd} Unit	**(Weight)**
1	2	**3**
1	4	**1**
2	3	**5**
2	4	**2**
2	5	**7**
4	5	**3**

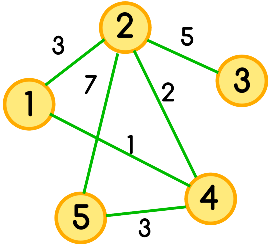
Classical network data

1^{st} Unit	2^{nd} Unit	(Weight)
1	2	**3**
1	4	**1**
2	3	**5**
2	4	**2**
2	5	**7**
4	5	**3**



Classical network data

1^{st} Unit	2^{nd} Unit	(Weight)
1	2	**3**
1	4	**1**
2	3	**5**
2	4	**2**
2	5	**7**
4	5	**3**



Classical graph metrics

- Connectedness and components

Classical graph metrics

- Connectedness and components
- Distance, average path length, clustering, efficiency

Classical graph metrics

- Connectedness and components
- Distance, average path length, clustering, efficiency
- Node centrality

Classical graph metrics

- Connectedness and components
- Distance, average path length, clustering, efficiency
- Node centrality
- Community structure

Classical graph metrics

- Connectedness and components
- Distance, average path length, clustering, efficiency
- Node centrality
- Community structure

$\Longrightarrow$ Processes on networks

Classical graph metrics

- Connectedness and components
- Distance, average path length, clustering, efficiency
- Node centrality
- Community structure

$\Longrightarrow$ Processes on networks (percolation, communication, spreading, synchronisation, opinions, etc.)

Time-resolved data

1st Unit	2nd Unit	Start	Length
2	4	0	40
2	5	50	10
2	3	70	20
4	5	60	50
1	2	130	15
1	4	140	35
2	3	220	20

Time-resolved data

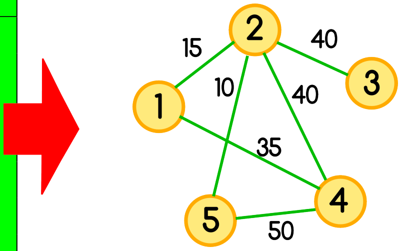
1^{st} Unit	2^{nd} Unit	Start	Length
2	4	0	40
2	5	50	10
2	3	70	20
4	5	60	50
1	2	130	15
1	4	140	35
2	3	220	20

Time-resolved data

1st Unit	2nd Unit	Start	Length
2	4	0	40
2	5	50	10
2	3	70	20
4	5	60	50
1	2	130	15
1	4	140	35
2	3	220	20

Time-resolved data

1^{st} Unit	2^{nd} Unit	Start	Length
2	4	0	40
2	5	50	10
2	3	70	20
4	5	60	50
1	2	130	15
1	4	140	35
2	3	220	20



Shortcomings of aggregated graphs

- Loss of temporal correlations and time-dependence

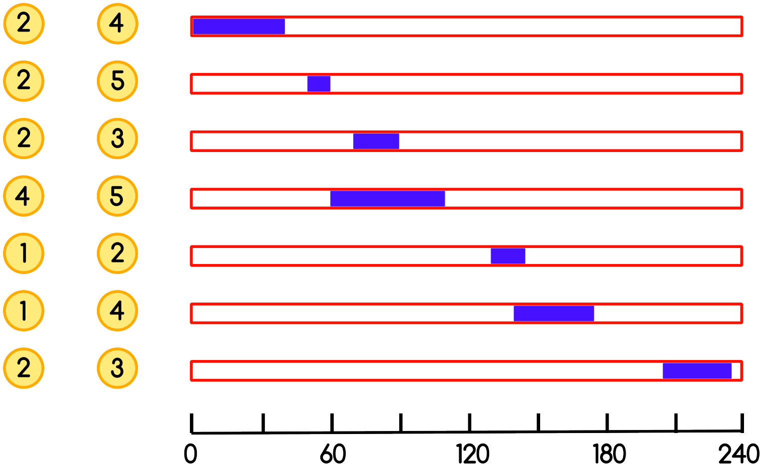
Shortcomings of aggregated graphs

- Loss of temporal correlations and time-dependence
- Overestimation of the number of available walks and paths

Adjacency: how does it change

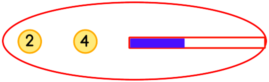
1st Unit	2nd Unit	Start	Length
2	4	0	40
2	5	50	10
2	3	70	20
4	5	60	50
1	2	130	15
1	4	140	35
2	3	220	20

Adjacency: how does it change



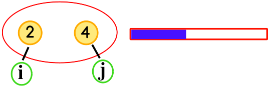
Contacts

- $c = (i, j, t, \delta t)$ is a contact
-
-
-



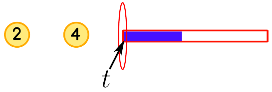
Contacts

- $c = (i, j, t, \delta t)$ is a contact
- i and j are two nodes
-
-



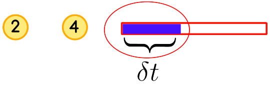
Contacts

- $c = (i, j, t, \delta t)$ is a contact
- i and j are two nodes
- t is the start time
-



Contacts

- $c = (i, j, t, \delta t)$ is a contact
- i and j are two nodes
- t is the start time
- δt is the contact duration



1st Unit	2nd Unit	(Weight)
1	2	**3**
1	4	**1**
2	3	**5**
2	4	**2**
2	5	**7**
4	5	**3**

1^{st} Unit	2^{nd} Unit	(Weight)
1	2	**3**
1	4	**1**
2	3	**5**
2	4	**2**
2	5	**7**
4	5	**3**

1^{st} Unit	2^{nd} Unit	(Weight)
1	2	**3**
1	4	**1**
2	3	**5**
2	4	**2**
2	5	**7**
4	5	**3**

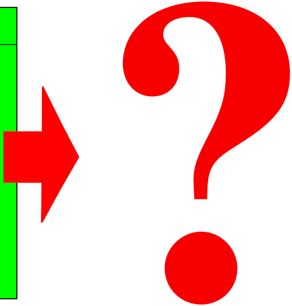
$$\Rightarrow$$

$$\begin{bmatrix} 0 & 3 & 0 & 1 & 0 \\ 3 & 0 & 5 & 2 & 7 \\ 0 & 5 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 3 \\ 0 & 7 & 0 & 3 & 0 \end{bmatrix}$$

1st Unit	2nd Unit	Start	Length
2	4	0	40
2	5	50	10
2	3	70	20
4	5	60	50
1	2	130	15
1	4	140	35
2	3	220	20

1^{st} Unit	2^{nd} Unit	Start	Length
2	4	0	40
2	5	50	10
2	3	70	20
4	5	60	50
1	2	130	15
1	4	140	35
2	3	220	20

1^{st} Unit	2^{nd} Unit	Start	Length
2	4	0	40
2	5	50	10
2	3	70	20
4	5	60	50
1	2	130	15
1	4	140	35
2	3	220	20



TVG: A formal definition

- Choose a time-window of size Δt

TVG: A formal definition

- Choose a time-window of size Δt
- $[t, t + \Delta t] \Longrightarrow G_t$ contains all the contacts $(\cdot, \cdot, \tau_i, \delta\tau_i)$ overlapping with $[t, t + \Delta t]$, i.e. such that:

$$t \leq \tau_i < t + \Delta t \quad \text{or} \tag{1}$$

$$t \leq \tau_i + \delta\tau_i < t + \Delta t \quad \text{or} \tag{2}$$

$$\tau_i < t \quad \wedge \quad \tau_i + \delta\tau_i > t + \Delta t \tag{3}$$

TVG: A formal definition

- Choose a time-window of size Δt
- $[t, t+\Delta t] \Longrightarrow G_t$ contains all the contacts $(\cdot, \cdot, \tau_i, \delta\tau_i)$ overlapping with $[t, t+\Delta t]$, i.e. such that:

$$
\begin{aligned}
t \le \tau_i < t + \Delta t \quad \text{or} \qquad (1)\\
t \le \tau_i + \delta\tau_i < t + \Delta t \quad \text{or} \qquad (2)\\
\tau_i < t \quad \wedge \quad \tau_i + \delta\tau_i > t + \Delta t \qquad (3)
\end{aligned}
$$

- G_t is a snapshot of the system in $[t, t+\Delta t]$.

TVG: A formal definition

- Choose a time-window of size Δt
- $[t, t + \Delta t] \Longrightarrow G_t$ contains all the contacts $(\cdot, \cdot, \tau_i, \delta\tau_i)$ overlapping with $[t, t + \Delta t]$, i.e. such that:

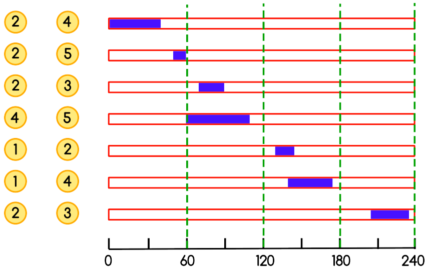
$$t \leq \tau_i < t + \Delta t \quad \text{or} \tag{1}$$

$$t \leq \tau_i + \delta\tau_i < t + \Delta t \quad \text{or} \tag{2}$$

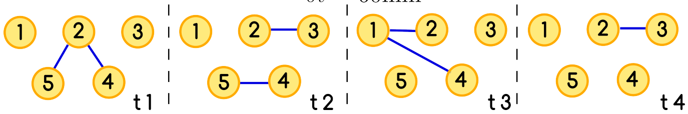
$$\tau_i < t \quad \wedge \quad \tau_i + \delta\tau_i > t + \Delta t \tag{3}$$

- G_t is a snapshot of the system in $[t, t + \Delta t]$.
- The sequence $\mathcal{G}_{0,T} = \{G_0, G_{\Delta t}, \dots G_T\}$ of M snapshots over N nodes is a time-varying graph.

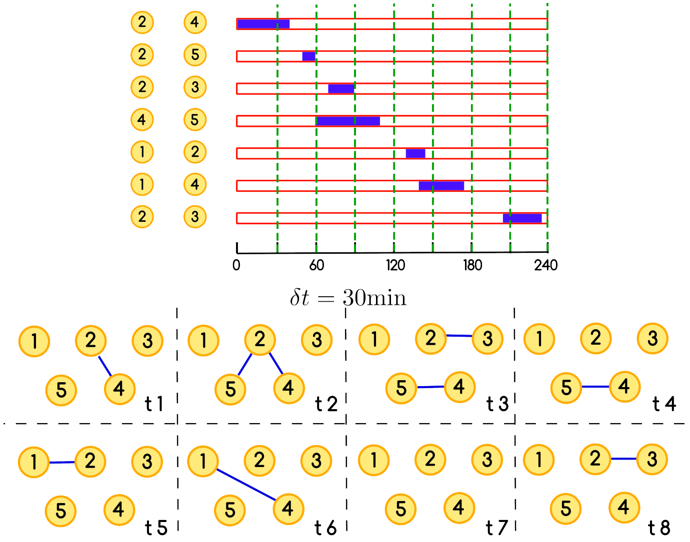
Time scales (1)



$\delta t = 60\text{min}$

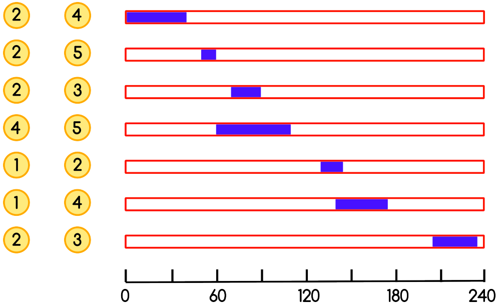


Time scales (2)



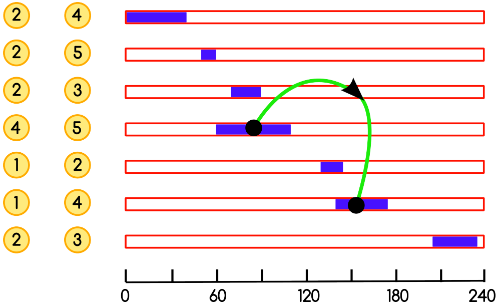
Reachability

From node 5 to node 1



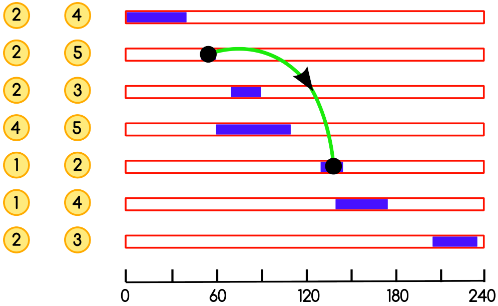
Reachability

From node 5 to node 1



Reachability

From node 5 to node 1



Reachability in TVGs

- A temporal walk from i to j is a temporally ordered sequence of L edges leading from i to j.

Reachability in TVGs

- A temporal walk from i to j is a temporally ordered sequence of L edges leading from i to j.
- A temporal path is a temporal walk for which each node is traversed exactly once.

Reachability in TVGs

- A temporal walk from i to j is a temporally ordered sequence of L edges leading from i to j.
- A temporal path is a temporal walk for which each node is traversed exactly once.
- The node i is temporally connected to j if there exists a temporal path leading from i to j

Reachability in TVGs

- A temporal walk from i to j is a temporally ordered sequence of L edges leading from i to j.
- A temporal path is a temporal walk for which each node is traversed exactly once.
- The node i is temporally connected to j if there exists a temporal path leading from i to j
- In this case, we say that j is temporally reachable from i.

Reachability in TVGs

- A temporal walk from i to j is a temporally ordered sequence of L edges leading from i to j.
- A temporal path is a temporal walk for which each node is traversed exactly once.
- The node i is temporally connected to j if there exists a temporal path leading from i to j
- In this case, we say that j is temporally reachable from i.
- Temporal connectedness IS NEITHER symmetric NOR transitive.

Node components

Given a node i we define:

- the temporal OUT-component of i (nodes j for which there is a TW from i to j)

Node components

Given a node i we define:

- the temporal OUT-component of i (nodes j for which there is a TW from i to j)
- the temporal IN-component of i (nodes j for which there is a TW from j to i)

Node components

Given a node i we define:

- the temporal OUT-component of i (nodes j for which there is a TW from i to j)
- the temporal IN-component of i (nodes j for which there is a TW from j to i)
- the temporal strongly connected component of i (nodes j which are both in $\mathrm{IN}(i)$ and in $\mathrm{OUT}(i)$)

Node components

Given a node i we define:

- the temporal OUT-component of i (nodes j for which there is a TW from i to j)
- the temporal IN-component of i (nodes j for which there is a TW from j to i)
- the temporal strongly connected component of i (nodes j which are both in $\mathrm{IN}(i)$ and in $\mathrm{OUT}(i)$)
- i and j are strongly connected if $i \in \mathrm{IN}(j)$ and $i \in \mathrm{OUT}(j)$

Graph components

- strongly connected component: a non-empty set of nodes S such that $\forall i, j \in S$ i and j are strongly connected

Graph components

- strongly connected component: a non-empty set of nodes S such that $\forall i, j \in S$ i and j are strongly connected
- Affine graph: a static graph $G_{\mathcal{G}}$ having the same nodes of $\mathcal{G}$ and such that (i, j) is an edge of $G_{\mathcal{G}}$ if i and j are strongly connected in $\mathcal{G}$

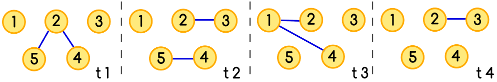
Graph components

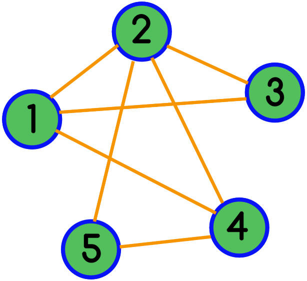
- strongly connected component: a non-empty set of nodes S such that $\forall i, j \in S$ i and j are strongly connected
- Affine graph: a static graph $G_{\mathcal{G}}$ having the same nodes of $\mathcal{G}$ and such that (i, j) is an edge of $G_{\mathcal{G}}$ if i and j are strongly connected in $\mathcal{G}$
- The strongly connected components of G are the maximal-cliques of $G_{\mathcal{G}}$

Graph components

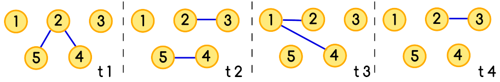
- strongly connected component: a non-empty set of nodes S such that $\forall i, j \in S$ i and j are strongly connected
- Affine graph: a static graph $G_{\mathcal{G}}$ having the same nodes of $\mathcal{G}$ and such that (i, j) is an edge of $G_{\mathcal{G}}$ if i and j are strongly connected in $\mathcal{G}$
- The strongly connected components of G are the maximal-cliques of $G_{\mathcal{G}}$
- Finding the largest strongly connected component of a TVG takes exponential time in the number of edges of the affine graph!

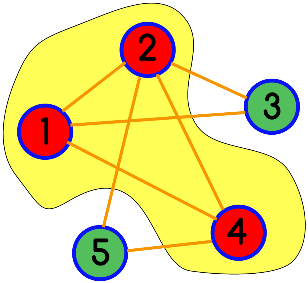
Affine graphs



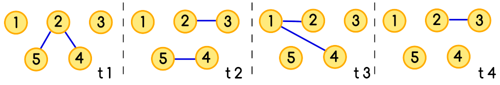


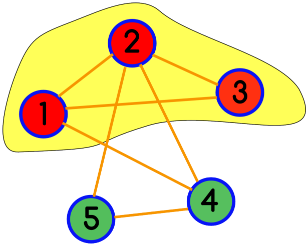
Affine graphs





Affine graphs





Application: Facebook

$\sim$ 100.000 profiles in Santa Barbara (CA) (2009)
1 week of messages

- Friendship network (static graph)
- Communication network (TVG – $\Delta t = 1$ hour)

Week	K	S	C
1	43491	22	12000
2	48404	20	13998
3	43400	16	12773
4	60853	41	17933
5	65703	23	19973
6	70282	27	20976
7	60666	28	18537
8	73772	46	20256
9	79645	38	21990
10	66849	18	20425
11	55040	27	18266
12	51418	28	15667

Lengths and distances

A time-respecting path has many different "lengths", namely:

- a topological length: the number of edges traversed by the path

Lengths and distances

A time-respecting path has many different "lengths", namely:

- a topological length: the number of edges traversed by the path
- temporal length or duration: the time interval between the first and the last contact in the path.

Lengths and distances

A time-respecting path has many different "lengths", namely:

- a topological length: the number of edges traversed by the path
- temporal length or duration: the time interval between the first and the last contact in the path.

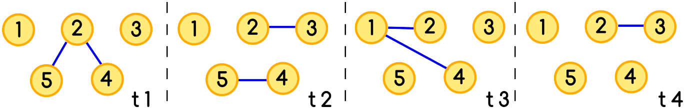
temporal shortest path: the temporal path connecting two nodes having minimum temporal length.

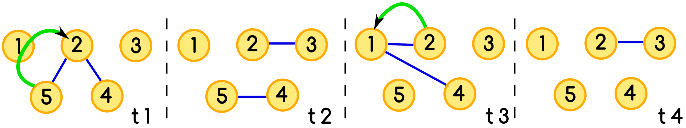
Lengths and distances

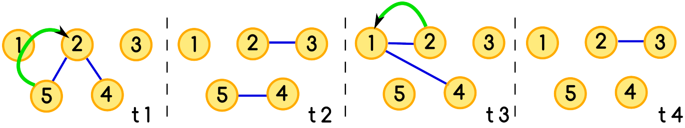
A time-respecting path has many different "lengths", namely:

- a topological length: the number of edges traversed by the path
- temporal length or duration: the time interval between the first and the last contact in the path.

temporal shortest path: the temporal path connecting two nodes having minimum temporal length.
temporal distance $d_{i,j}$ is the temporal length of the temporal shortest path from i to j.


1
2
3
5
4
t 1
1
2
3
5
4
t 2
1
2
3
5
4
t 3
1
2
3
5
4
t 4


1
2
3
5
4
t 1
1
2
3
5
4
t 2
1
2
3
5
4
t 3
1
2
3
5
4
t 4



- Topological length: 2
- Temporal length: $3\Delta t$

Length-related metrics

Average temporal length

$$L = \frac{1}{N(N-1)} \sum_{ij} d_{ij} \tag{4}$$

Length-related metrics

Average temporal length

$$L = \frac{1}{N(N-1)} \sum_{ij} d_{ij} \tag{4}$$

Temporal diamater:

$$D = max_{ij} d_{ij} \tag{5}$$

Length-related metrics

Average temporal length

$$L = \frac{1}{N(N-1)} \sum_{ij} d_{ij} \tag{4}$$

Temporal diamater:

$$D = max_{ij} d_{ij} \tag{5}$$

Temporal efficiency:

$$\mathcal{E} = \frac{1}{N(N-1)} \sum_{ij} \frac{1}{d_{ij}} \tag{6}$$

Application: node percolation

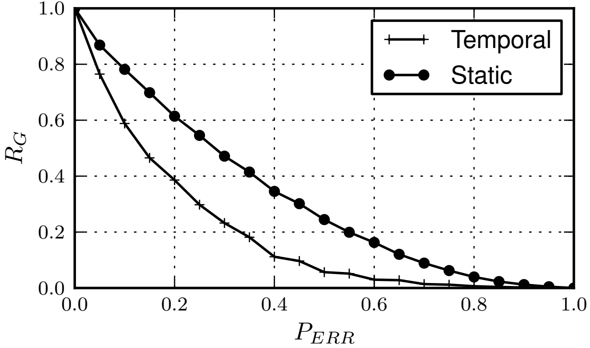
- Damage: $D\%$ of the nodes are removed (percolated) from the network $\Longrightarrow$ new graph $\mathcal{G}_{\mathcal{D}}$

Application: node percolation

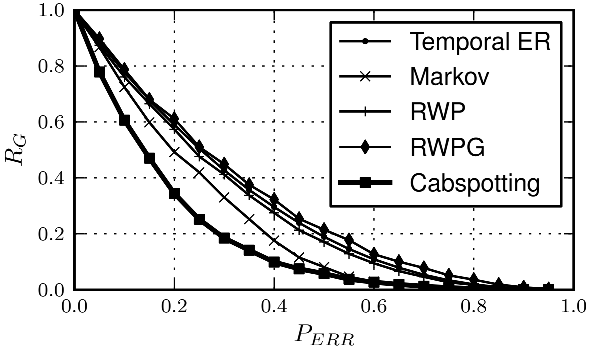
- Damage: $D\%$ of the nodes are removed (percolated) from the network $\Longrightarrow$ new graph $\mathcal{G}_{\mathcal{D}}$
- Robustness:

$$R = \frac{E_{\mathcal{G}_D}}{E_{\mathcal{G}}} \tag{7}$$

Cabspotting: aggregated vs TVG



Cabspotting: TVG vs random models



Temporal Clustering

Topological overlap of the neighbourhood of i in $[t_m, t_{m+1}]$:

$$C_i(t_m, t_{m+1}) = \frac{\sum_j a_{ij}(t_m) a_{ij}(t_{m+1})}{\sqrt{\left[\sum_j a_{ij}(t_m)\right]\left[\sum_j a_{ij}(t_{m+1})\right]}} \tag{8}$$

Temporal Clustering

Topological overlap of the neighbourhood of i in $[t_m, t_{m+1}]$:

$$C_i(t_m, t_{m+1}) = \frac{\sum_j a_{ij}(t_m) a_{ij}(t_{m+1})}{\sqrt{\left[\sum_j a_{ij}(t_m)\right]\left[\sum_j a_{ij}(t_{m+1})\right]}} \tag{8}$$

Average topological overlap:

$$C_i = \frac{1}{M-1} \sum_{m=1}^{M-1} C_i(t_m, t_{m+1}) \tag{9}$$

Temporal Clustering

Topological overlap of the neighbourhood of i in $[t_m, t_{m+1}]$:

$$C_i(t_m, t_{m+1}) = \frac{\sum_j a_{ij}(t_m) a_{ij}(t_{m+1})}{\sqrt{\left[\sum_j a_{ij}(t_m)\right]\left[\sum_j a_{ij}(t_{m+1})\right]}} \tag{8}$$

Average topological overlap:

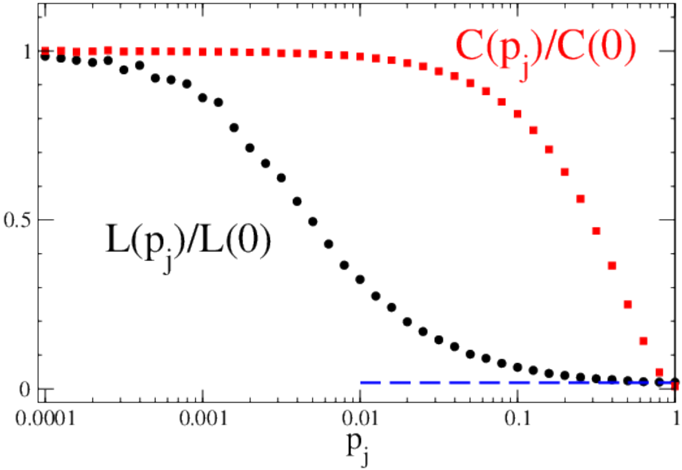
$$C_i = \frac{1}{M-1} \sum_{m=1}^{M-1} C_i(t_m, t_{m+1}) \tag{9}$$

Temporal correlation coefficient

$$C = \frac{1}{N} \sum_i C_i \tag{10}$$

Temporal small-world effect

	C	C^{rand}	L	L^{rand}	E	E^{rand}
α	0.44	0.18 (0.03)	3.9	4.2	0.50	0.48
β	0.40	0.17 (0.002)	6.0	3.6	0.41	0.45
γ	0.48	0.13 (0.003)	12.2	8.7	0.39	0.37
δ	0.44	0.17 (0.003)	2.2	2.4	0.57	0.56
d1	0.80	0.44 (0.01)	8.84	6.00	0.192	0.209
d2	0.78	0.35 (0.01)	5.04	4.01	0.293	0.298
d3	0.81	0.38 (0.01)	9.06	6.76	0.134	0.141
d4	0.83	0.39 (0.01)	21.42	15.55	0.019	0.028
Mar	0.044	0.007 (0.0002)	456	451	0.000183	0.000210
Jun	0.046	0.006 (0.0002)	380	361	0.000047	0.000057
Sep	0.046	0.006 (0.0002)	414	415	0.000058	0.000074
Dec	0.049	0.006 (0.0002)	403	395	0.000047	0.000059


C(p_j)/C(0)
L(p_j)/L(0)
1
0.5
0
0.0001
0.001
0.01
0.1
1
p_j

Betweenness and closeness centrality

Temporal betweenness centrality of a node at time t_m:

$$C_i^B(t_m) = \frac{1}{(N-1)(N-2)} \sum_{j \neq i} \sum_{\substack{k \neq j \\ k \neq i}} \frac{U(i, t_m, j, k)}{\sigma_{jk}} \tag{11}$$

Betweenness and closeness centrality

Temporal betweenness centrality of a node at time t_m:

$$C_i^B(t_m) = \frac{1}{(N-1)(N-2)} \sum_{j \neq i} \sum_{\substack{k \neq j \\ k \neq i}} \frac{U(i, t_m, j, k)}{\sigma_{jk}} \tag{11}$$

Average temporal betweenness of node i:

$$C_i^B = \frac{1}{M} \sum_m C_i^B(t_m) \tag{12}$$

Betweenness and closeness centrality

Temporal betweenness centrality of a node at time t_m:

$$C_i^B(t_m) = \frac{1}{(N-1)(N-2)} \sum_{j \neq i} \sum_{\substack{k \neq j \\ k \neq i}} \frac{U(i, t_m, j, k)}{\sigma_{jk}} \tag{11}$$

Average temporal betweenness of node i:

$$C_i^B = \frac{1}{M} \sum_m C_i^B(t_m) \tag{12}$$

Average temporal closeness of i:

$$C_i^C = \frac{N-1}{\sum_j d_{ij}} \tag{13}$$

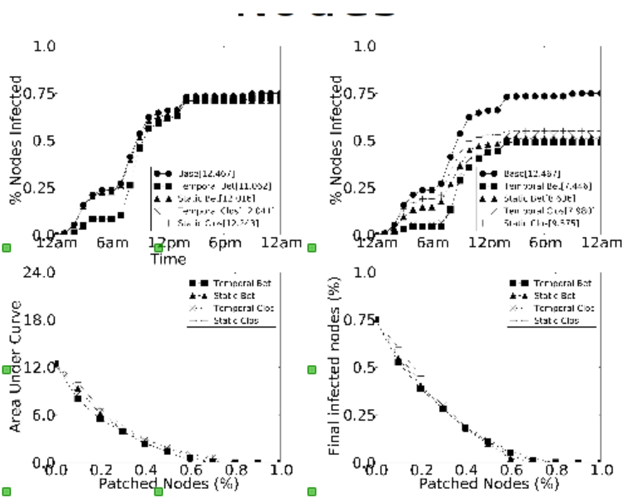
Application: information spreading & success

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

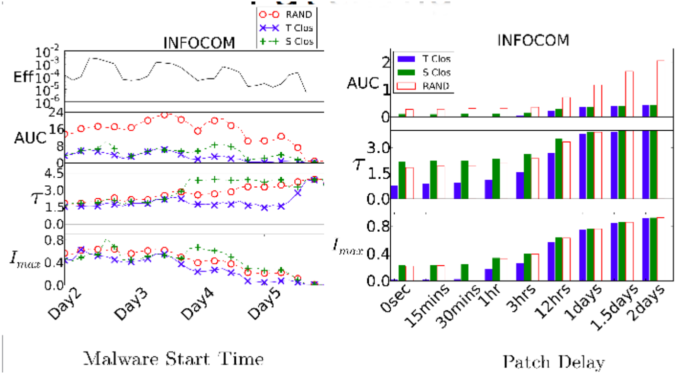
Top bonuses awarded

John Lavorato: $5 million
Louise Kitchen: $2 million
Jeffrey McMahon: $1.5 million
James Fallon: $1.5 million
Raymond Bowen Jr.: $750,000
Mark Haedicke: $750,000
Gary Hickerson: $700,000
Wesley Colwell: $600,000
Richard Dimichele: $600,000

Application: mobile malware containment



Application: mobile malware containment



References

- Tang, J., Scellato, S., Musolesi, M., Mascolo, C., and Latora, V. *Small World behavior in temporal networks* Phys. Rev. E 81,, 055101(2010).
- Nicosia, V., Tang, J., Musolesi, M., Russo, G., Mascolo, C., and Latora, V. *Components in time-varying graphs.* Chaos, 22,, 023101.
- Tang, J., Musolesi, M., Mascolo, C., and Latora, V. SIGCOMM Comput. Com- mun. Rev. 40(1), 118124.
- J. Tang, M. Musolesi, C. Mascolo, and V. Latora. *Temporal Distance Metrics for Social Network Analysis.* In Proceedings of WOSN '09, Barcelona, Spain, August 2009.
- S. Scellato, I. Leontiadis, C. Mascolo, P. Basu, and M. Zafer. *Evaluating Temporal Robustness of Mobile Networks.* IEEE Transactions on Mobile Computing.
- J. Tang, M. Musolesi, C. Mascolo, V. Latora, and V. Nicosia. *Analysing information flows and key mediators through temporal centrality metrics.* In Proceedings of ACM SNS '10, Paris, France, April 2010.
- J. Tang, H. Kim, C. Mascolo, M. Musolesi. *STOP: Socio-Temporal Opportunistic Patching of Short Range Mobile Malware.* In Proceedings of the 13th IEEE Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM'12).
- J. Tang, C. Mascolo, M. Musolesi, and V. Latora. *Exploiting Temporal Complex Network Metrics in Mobile Malware Containment.* In Proceedings of the 12th IEEE Interna- tional Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM'11),