Temporal Distance Metrics for Social Networks Analysis

John Tang¹, Mirco Musolesi¹, Cecilia Mascolo¹ and Vito Latora²

Computer Laboratory, University of Cambridge
²INFN/Dept of Physics, University of Catania
Hyperbolic view of BGP tables

Credit: kc claffy, CAIDA

mirco.musolesi@cl.cam.ac.uk
Problem: existing metrics do not capture the inherent dynamism of networks over time.

We need new **temporal metrics** defined over **temporal graphs** for studying dynamic processes over these networks.
An Example of Temporal Graph

A  B
C  D
E  F

A  B
C  D
E  F

A  B
C  D
E  F

A  B
C  D
E  F

A  B
C  D
E  F

A  B
C  D
E  F

mirmo.musolesi@cl.cam.ac.uk
...and the Corresponding Static Graph
Calculating the Temporal Distance

A

B

C

D

E

F

t = 1

t = 2

t = 3

t = 4

mirco.musolesi@cl.cam.ac.uk
Calculating the Temporal Distance

\((t = 1)\)

D at distance 1

A

B

C

D

E

F

A

B

C

D

E

F

t = 1

t = 2

t = 3

t = 4

mirco.musolesi@cl.cam.ac.uk
Calculating the Temporal Distance
(t = 2)
Calculating the Temporal Distance
(t = 3)

B and C at distance 3

A
B
C
D
E
F

A
B
C
D
E
F

A
B
C
D
E
F

A
B
C
D
E
F

A
B
C
D
E
F

mirco.musolesi@cl.cam.ac.uk
Calculating the Temporal Distance
(t = 4)

A B A B A B A B
C D C D C D C D
E F E F E F E F

F at distance 4

t = 1 t = 2 t = 3 t = 4

mirco.musolesi@cl.cam.ac.uk
What about the Static Distance?

E is *statically* reachable but in reality it is not *dynamically* reachable!

A→F requires 2 transmissions (hops), but in reality it requires 3

No information about the duration of the process
What about the Symmetric Distance (F to A)?

mirco.musolesi@cl.cam.ac.uk
Calculating the Inverse Temporal Distance

\( t = 1 \)

\[ A \quad B \quad A \quad B \quad A \quad B \quad A \quad B \]

\[ C \quad D \quad C \quad D \quad C \quad D \quad C \quad D \]

\[ E \quad F \quad E \quad F \quad E \quad F \quad E \quad F \]

\[ t = 1 \quad t = 2 \quad t = 3 \quad t = 4 \]

mirco.musolesi@cl.cam.ac.uk
Calculating the Inverse Temporal Distance
(t = 2)
Calculating the Inverse Temporal Distance
\( (t = 3) \)
Calculating the Inverse Temporal Distance 
(t = 4)

A is not reachable [infinite distance]
Let’s Get a Bit More Formal...

• Characteristic temporal path length:

\[ L^h(t_{min}, t_{max}) = \frac{1}{N(N - 1)} \sum_{ij} d_{ij}^h(t_{min}, t_{max}) \]

• Defined considering the horizon of the infection

• Possible problem due to the potential divergence due to pairs of nodes that are not temporally connected
Let’s Get a Bit More Formal...

• Characteristic temporal path length:

\[ L^h(t_{min}, t_{max}) = \frac{1}{N(N - 1)} \sum_{ij} d^h_{ij}(t_{min}, t_{max}) \]

• Defined considering the horizon of the infection

• Possible problem due to the potential divergence due to pairs of nodes that are not temporally connected
Impact of the Horizon Parameter

(F -> A, h = 1)
Impact of the Horizon Parameter
\((F \rightarrow A, h = 2)\)

A was not reachable at all with \(h = 1\) (in 4 time windows), but with \(h = 2\) it is a distance 1!

\(t = 1\)  \(t = 2\)  \(t = 3\)  \(t = 4\)

mirco.musolesi@cl.cam.ac.uk
Let’s Get a Bit More Formal…

- Characteristic temporal path length:

\[ L^h(t_{\text{min}}, t_{\text{max}}) = \frac{1}{N(N-1)} \sum_{i,j} d^h_{ij}(t_{\text{min}}, t_{\text{max}}) \]

- Defined considering the horizon of the infection

- Possible problem related to the potential divergence due to pairs of nodes that are not temporally connected
Let’s Get a Bit More Formal...

• Solution: definition of temporal efficiency:

\[
E_{T_{ij}}^h(t_{min}, t_{max}) = \frac{1}{d_{ij}^h(t_{min}, t_{max})}
\]

\[
E_{glob}^h(t_{min}, t_{max}) = \frac{1}{N(N-1)} \sum_{ij} E_{T_{ij}}^h(t_{min}, t_{max})
\]

• High value of E (low value of L) means that the nodes of the graphs can communicate efficiently
Local Temporal Efficiency

\[ E_{loc_i}(t_{min}, t_{max}) = E_T \{ G_i^{N_i(t_{min}, t_{max})} \quad t \in [t_{min}, t_{max}] \} \]

\[ E_{loc}(t_{min}, t_{max}) = \frac{1}{N} \sum_i E_{loc_i}(t_{min}, t_{max}) \]
Temporal Clustering Coefficient

\[ C_i(t_{\text{min}}, t_{\text{max}}) = \frac{\sum_{t=t_{\text{min}}}^{t_{\text{max}}} \# \text{ of edges in } G_t^{\mathcal{N}_i(t_{\text{min}}, t_{\text{max}})}}{\tau \cdot \frac{k_i(t_{\text{min}}, t_{\text{max}})[k_i(t_{\text{min}}, t_{\text{max}})-1]}{2}} \]

\[ C(t_{\text{min}}, t_{\text{max}}) = \frac{1}{N} \sum_i C_i(t_{\text{min}}, t_{\text{max}}) \]
### INFOCOM Dataset: Static vs Temporal Metrics

<table>
<thead>
<tr>
<th>Day</th>
<th>N</th>
<th>$\langle k \rangle$</th>
<th>Static</th>
<th></th>
<th></th>
<th>Temporal</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>C</td>
<td>L</td>
<td>$C_{rand}$</td>
<td>$L_{rand}$</td>
<td>C</td>
<td>$L^*$</td>
<td>Disc</td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>25.7</td>
<td>0.818</td>
<td>1.291</td>
<td>0.764</td>
<td>1.336</td>
<td>0.033</td>
<td>4.090</td>
<td>0.28</td>
</tr>
<tr>
<td>2</td>
<td>39</td>
<td>28.3</td>
<td>0.845</td>
<td>1.269</td>
<td>0.824</td>
<td>1.263</td>
<td>0.110</td>
<td>4.556</td>
<td>0.13</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
<td>22.3</td>
<td>0.744</td>
<td>1.420</td>
<td>0.644</td>
<td>1.405</td>
<td>0.077</td>
<td>4.003</td>
<td>0.19</td>
</tr>
<tr>
<td>4</td>
<td>39</td>
<td>21.4</td>
<td>0.722</td>
<td>1.444</td>
<td>0.541</td>
<td>1.474</td>
<td>0.052</td>
<td>4.705</td>
<td>0.14</td>
</tr>
</tbody>
</table>

$h_{min} = 12\text{am}, \ h_{max} = 12\text{pm}, \ w = 5\ \text{min}$

Static metrics underestimate $L$
### Reality Mining Dataset

<table>
<thead>
<tr>
<th>Date</th>
<th>C</th>
<th>Eloc</th>
<th>L</th>
<th>Eglob</th>
<th>Eloc</th>
<th>L</th>
<th>Eglob</th>
</tr>
</thead>
<tbody>
<tr>
<td>08 Sep</td>
<td>0.014</td>
<td>0.000</td>
<td>23h 15m</td>
<td>0.000</td>
<td>0.003</td>
<td>21h 58m</td>
<td>0.010</td>
</tr>
<tr>
<td>15 Sep</td>
<td>0.060</td>
<td>0.000</td>
<td>22h 47m</td>
<td>0.001</td>
<td>0.007</td>
<td>19h 55m</td>
<td>0.024</td>
</tr>
<tr>
<td>22 Sep</td>
<td>0.061</td>
<td>0.000</td>
<td>22h 53m</td>
<td>0.001</td>
<td>0.007</td>
<td>20h 42m</td>
<td>0.019</td>
</tr>
<tr>
<td>29 Sep</td>
<td>0.060</td>
<td>0.001</td>
<td>22h 20m</td>
<td>0.001</td>
<td>0.009</td>
<td>17h 44m</td>
<td>0.037</td>
</tr>
<tr>
<td>06 Oct</td>
<td>0.026</td>
<td>0.000</td>
<td>22h 14m</td>
<td>0.001</td>
<td>0.011</td>
<td>16h 23m</td>
<td>0.041</td>
</tr>
<tr>
<td>13 Oct</td>
<td>0.038</td>
<td>0.000</td>
<td>21h 37m</td>
<td>0.004</td>
<td>0.013</td>
<td>14h 57m</td>
<td>0.055</td>
</tr>
<tr>
<td>20 Oct</td>
<td>0.067</td>
<td>0.001</td>
<td>21h 45m</td>
<td>0.003</td>
<td>0.007</td>
<td>17h 4m</td>
<td>0.031</td>
</tr>
<tr>
<td>27 Oct</td>
<td>0.050</td>
<td>0.002</td>
<td>22h 1m</td>
<td>0.001</td>
<td>0.013</td>
<td>15h 19m</td>
<td>0.050</td>
</tr>
<tr>
<td>03 Nov</td>
<td>0.051</td>
<td>0.001</td>
<td>21h 6m</td>
<td>0.004</td>
<td>0.012</td>
<td>16h 17m</td>
<td>0.043</td>
</tr>
<tr>
<td>10 Nov</td>
<td>0.051</td>
<td>0.000</td>
<td>20h 5m</td>
<td>0.004</td>
<td>0.015</td>
<td>14h 25m</td>
<td>0.061</td>
</tr>
</tbody>
</table>

$h = 1$, $t_{\text{min}} = 12\text{ am}$, $t_{\text{max}} = 12\text{pm}$, $w = 5\text{ min}$
Variation of the Clustering Coefficient over Time

Reality Mining Dataset ($t_{\text{min}} = 12$ am, $t_{\text{max}} = 12$ pm, $w = 5$ min)
Temporal Efficiency vs Static Efficiency

(a) Graph showing the trend of \( E \) over the months from 1 to 9.

(b) Graph showing the trend of \( E_{\text{stat}} \) over the months from 1 to 9.

(c) Graph showing the trend of the number of contacts over the months from 1 to 9.

(d) Graph showing the trend of the number of active nodes over the months from 1 to 9.

MIT Dataset

mirco.musolesi@cl.cam.ac.uk
Current Research Agenda

• Centrality measures
• Study of dynamic processes over the networks:
  – Message dissemination
  – Epidemics
  – Information propagation
• Analysis of new and larger datasets
• Small-world behavior in temporal networks
Summary of the Talk

• New temporal metrics for studying dynamic processes over dynamic networks
  – Temporal distance
  – Temporal efficiency
  – Temporal clustering

• Analysis using real datasets

mirco.musolesi@cl.cam.ac.uk
Questions?

Mirco Musolesi

mirco.musolesi@cl.cam.ac.uk
[soon to be: mirco@cs.st-andrews.ac.uk]

http://www.cl.cam.ac.uk/~mm753
INFOCOM Dataset:
What Happens if We Reshuffle the Sequence?

<table>
<thead>
<tr>
<th>Day</th>
<th>C</th>
<th>$E_{loc}$</th>
<th>L</th>
<th>$E_{glob}$</th>
<th>$E_{loc}$</th>
<th>L</th>
<th>$E_{glob}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.033</td>
<td>0.003</td>
<td>19h 39m</td>
<td>0.003</td>
<td>0.077</td>
<td>5h 29m</td>
<td>0.100</td>
</tr>
<tr>
<td>2</td>
<td>0.110</td>
<td>0.020</td>
<td>9h 6m</td>
<td>0.024</td>
<td>0.194</td>
<td>2h 45m</td>
<td>0.239</td>
</tr>
<tr>
<td>3</td>
<td>0.077</td>
<td>0.013</td>
<td>10h 32m</td>
<td>0.018</td>
<td>0.114</td>
<td>4h 6m</td>
<td>0.167</td>
</tr>
<tr>
<td>4</td>
<td>0.052</td>
<td>0.009</td>
<td>9h 55m</td>
<td>0.013</td>
<td>0.104</td>
<td>3h 3m</td>
<td>0.165</td>
</tr>
</tbody>
</table>

h = 1, $t_{min}$ = 12 am, $t_{max}$ = 12 pm, no runs = 50

mirco.musolesi@cl.cam.ac.uk
### Email Dataset

**Temporal Metrics**

<table>
<thead>
<tr>
<th>Date</th>
<th>C</th>
<th>$E_{loc}$</th>
<th>L</th>
<th>$E_{glob}$</th>
<th>$E_{loc}$</th>
<th>L</th>
<th>$E_{glob}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>27Oct</td>
<td>0</td>
<td>$3.1E^{-8}$</td>
<td>86397.94s</td>
<td>$9.3E^{-7}$</td>
<td>$7.7E^{-8}$</td>
<td>86396.91s</td>
<td>$1.6E^{-6}$</td>
</tr>
<tr>
<td>28Oct</td>
<td>$3.5E^{-7}$</td>
<td>$4.0E^{-8}$</td>
<td>86399.78s</td>
<td>$1.4E^{-7}$</td>
<td>$4.1E^{-8}$</td>
<td>86399.71s</td>
<td>$1.5E^{-7}$</td>
</tr>
<tr>
<td>29Oct</td>
<td>$2.5E^{-7}$</td>
<td>$3.9E^{-8}$</td>
<td>86399.03s</td>
<td>$3.9E^{-7}$</td>
<td>$7.2E^{-8}$</td>
<td>86398.59s</td>
<td>$7.3E^{-7}$</td>
</tr>
<tr>
<td>30Oct</td>
<td>0</td>
<td>$5.8E^{-8}$</td>
<td>86398.76s</td>
<td>$5.5E^{-7}$</td>
<td>$6.9E^{-8}$</td>
<td>86398.48s</td>
<td>$7.5E^{-7}$</td>
</tr>
<tr>
<td>31Oct</td>
<td>0</td>
<td>$4.7E^{-8}$</td>
<td>86398.92s</td>
<td>$4.9E^{-7}$</td>
<td>$6.5E^{-8}$</td>
<td>86398.64s</td>
<td>$6.9E^{-7}$</td>
</tr>
<tr>
<td>01Nov</td>
<td>0</td>
<td>$5.8E^{-8}$</td>
<td>86399.03s</td>
<td>$4.9E^{-7}$</td>
<td>$6.6E^{-8}$</td>
<td>86398.85s</td>
<td>$6.0E^{-7}$</td>
</tr>
<tr>
<td>02Nov</td>
<td>0</td>
<td>$4.3E^{-8}$</td>
<td>86398.68s</td>
<td>$5.4E^{-7}$</td>
<td>$6.5E^{-8}$</td>
<td>86398.67s</td>
<td>$6.8E^{-7}$</td>
</tr>
</tbody>
</table>

$h = 1$, $t_{\text{min}} = 12 \text{ am}$, $t_{\text{max}} = 12 \text{ pm}$, $w = 5 \text{ min}$