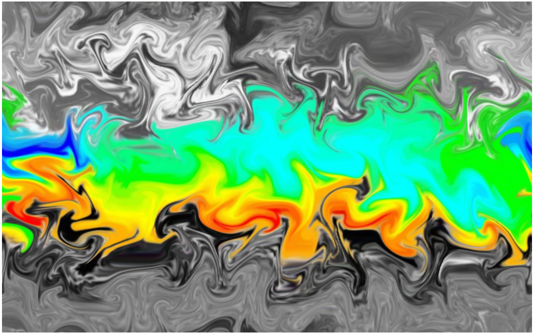
Modeling and Measuring Graph Similarity: The Case for Centrality Distance



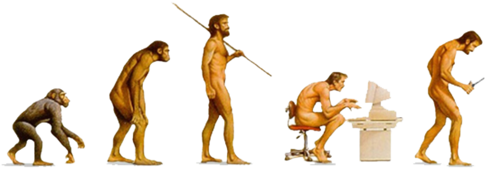
Theoretical and Computational Fluid Dynamics Laboratory – Blair Perot

Matthieu Roy[1], Stefan Schmid[2], *Gilles Tredan*[1]

1– Laboratory for Analysis and Architecture of Systems
2– TU-Berlin/Deutsche Telekom

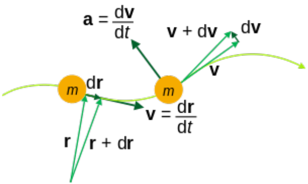
Dynamic Graphs

- Everywhere
- Huge
- We can't measure them easily



Distance

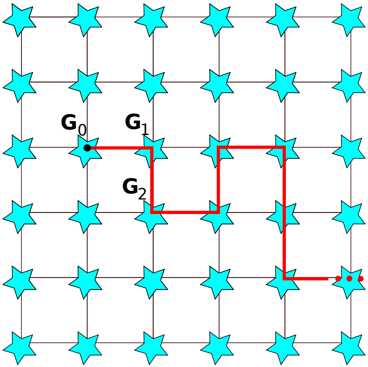
- =Root of Dynamism Characterisation
- Interpolation
- Extrapolation
- Coordinate System,....



Which distances for graphs ?

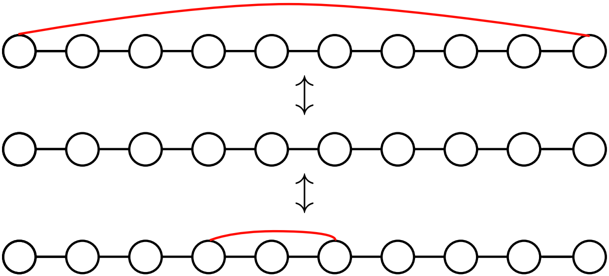
Model

- Named Graphs:
  Alice–Eve–Bob ! = Alice–Bob–Eve
- Undirected graphs (should be ok for directed ones)
- All nodes present at $t = 0$
- "Dynamicity": stream of edge additions/deletions:
  $G_0 = (V, E_0), G_1 = (V, E_1), \ldots$



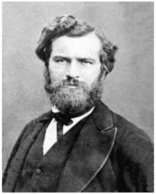
Graph Edit Distance

- Graph Edit Distance = d_{GED}
- **Only** known proper graph distance
- $d_{GED}(A, B)$ = number of graph edit operations from A to B
- Named graphs $\Leftrightarrow$ Cheap
- Too "blunt"



Centralities

- Borgatti,Everett (2006): "the only thing people agree about a centrality is that it is a node-level measure".
- Loved by SNAnalysts



Camille Jordan

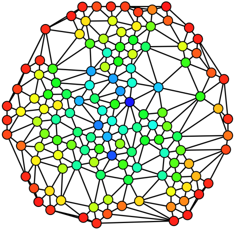
Definition (Centrality)

A *centrality* C is a function $C: (G, v) \to \mathbb{R}^+$ that takes a graph $G = (V, E)$ and a vertex $v \in V(G)$ and returns a positive value $C(G, v)$.

Degree $c_d(G, i) = degree(i)$

Closeness $c_c(G, i) = \sum_{j \neq i} d_G(i, j)$

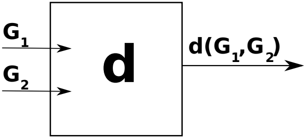
Betweenness $c_c(G, i) = \sum_{j \neq i, k \neq i} \delta_{i \in sp(j,k)}$

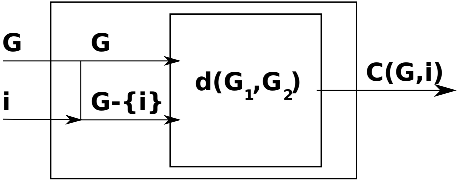


"Graph betweenness" - Claudio Rocchini

What About Other Distances ?

- Imagine a graph distance $d : \mathcal{G}^2 \rightarrow \mathbb{R}^+$
- I can construct you a centrality out of it ! $c^d(G, v) = d(G, G - \{i\})$





Very intriguing: $d_{GED}(G, G - \{i\}) = degree_G(i)$

A Connection Between Both Concepts ?

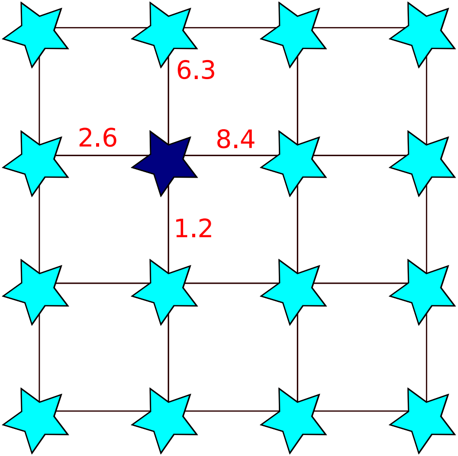
- We have many centralities, few distances.
- Can we construct the other way round ? Yes!

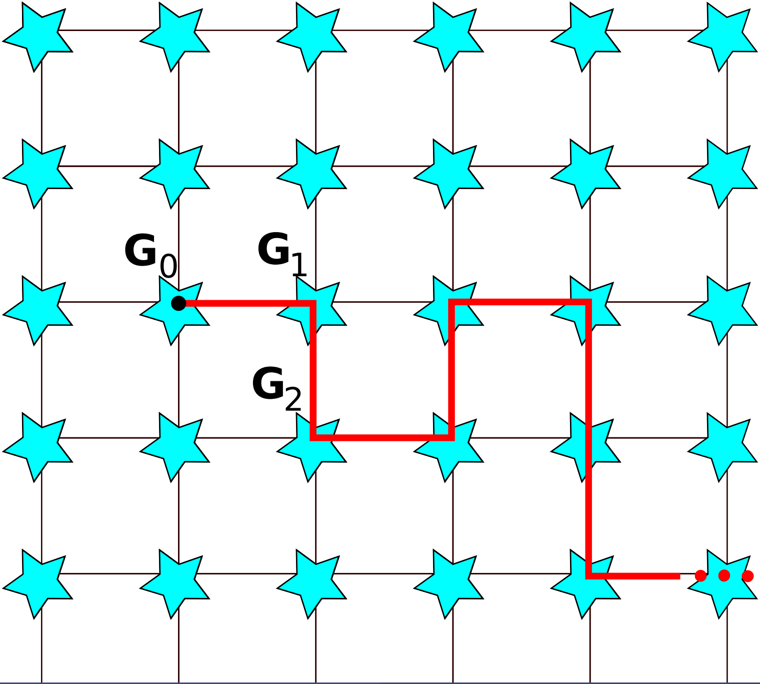
Definition (Centrality Distance)

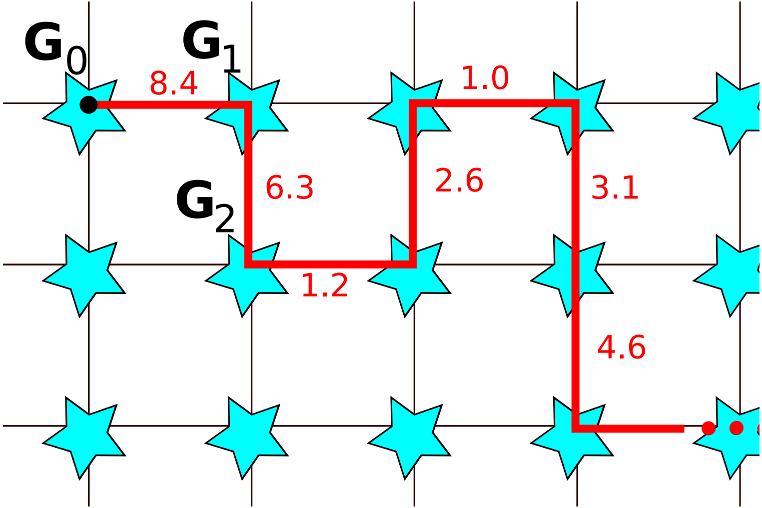
Given a centrality C, we define the centrality distance $d_C(G_1, G_2)$ between two *neighboring* graphs as the component-wise difference:

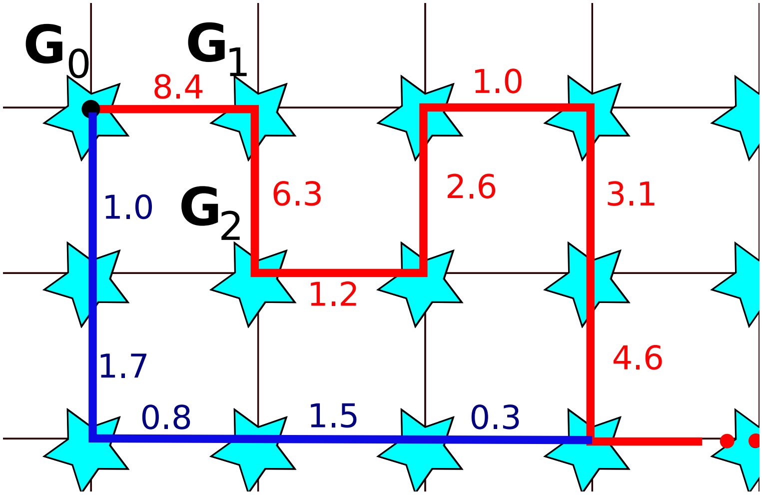
$$\forall (G_1, G_2) \in E(\mathcal{G}), d_C(G_1, G_2) = \sum_{v \in V} |C(G_1, v) - C(G_2, v)|.$$

Natural extension for *non-neighboring* graph couples: $d_C(G_1, G_2)$ = graph-induced distance on the valued graph $\mathcal{G}$.


6.3
2.6
8.4
1.2


G_0
G_1
G_2


G_0
G_1
G_2
8.4
1.0
6.3
2.6
3.1
1.2
4.6



G0
G1
G2
8.4
1.0
6.3
2.6
3.1
1.2
4.6
1.0
1.7
0.8
1.5
0.3

Connection Cont.

Definition (Sensitive Centrality)

Centrality C is *sensitive* iff

$$\forall G \in \mathcal{G}, \forall e \in E(G), \exists v \in V(G) \text{ s.t. } C(G,v) \neq C(G \setminus \{e\}, v),$$

d_C is a distance *iff* C is sensitive

- Not all sensitive ! Ex: Excentricity
- Some centralities need adaptations
- Approximate (cheap) version:

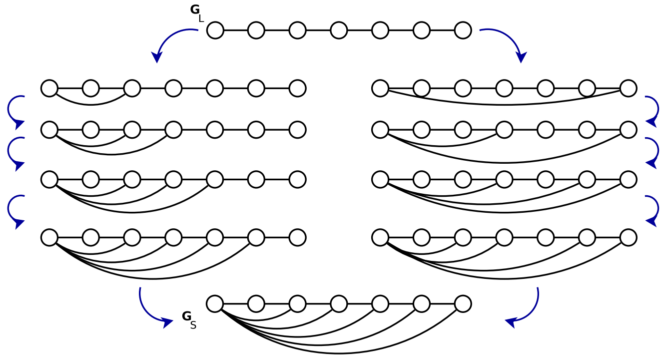
$$\forall (G_1, G_2), \quad \widetilde{d_C}(G_1, G_2) = \sum_{v \in V} |C(G_1, v) - C_2(G_2, v)|.$$

- co-central graphs!

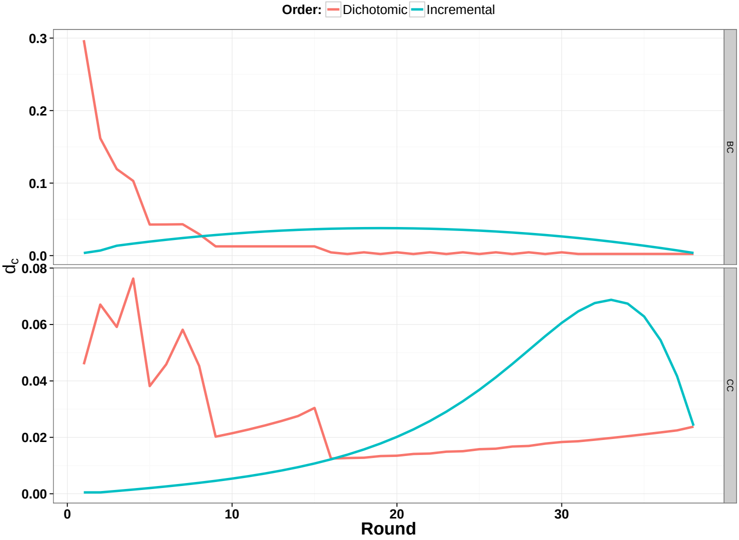
Experiments

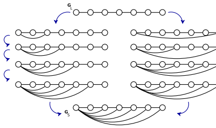
Topology evolution

- Differentiate between paths using centrality-induced distances
- These paths are equivalent wrt d_{GED}
- Yet they wouldn't impact networks the same way.



Topology evolution Cont.

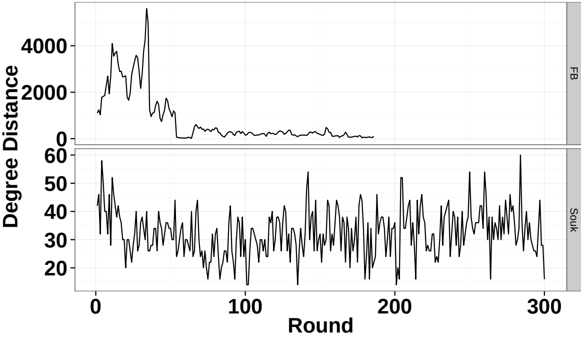




Dynamic Topologies

2 Datasets

- Facebook like OSN
  - Online messages exchange
  - ≈ 20k users
  - 187 snapshots, 1 day sampling
- Souk mobility dataset
  - Social contacts within a crowd
  - 45 individuals
  - 300 snapshots, 3 sec. sampling

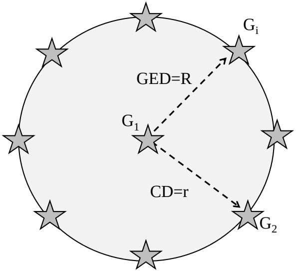


Dynamic Topologies

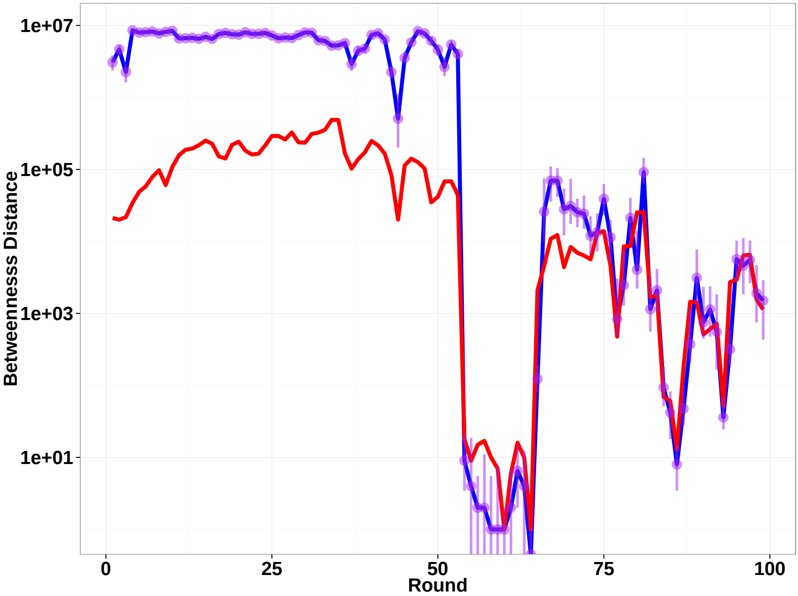
Can we distinguish "natural" evolutions from artificial ones ?

Methodology

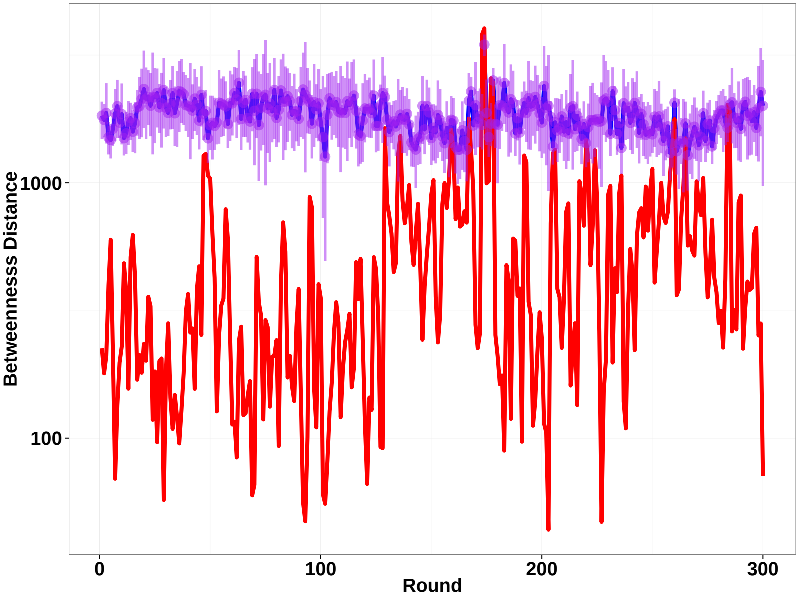
- For every two consecutive snapshots G_t, G_{t+1}
- Generate 200 random graphs $G_{s_1}, \ldots. G_{s_{200}}$
- **with** $d_{GED}(G_t, G_{t+1}) = d_{GED}(G_t, G_{s_i})$
- Compare $d_C(G_t, G_{t+1})$ with $d_C(G_t, G_{s_i})$



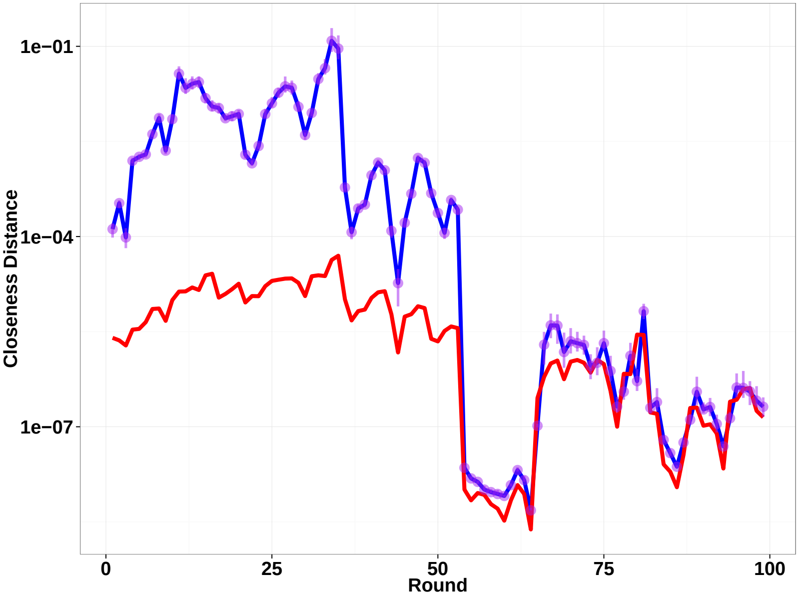
Dynamic Topologies/Results



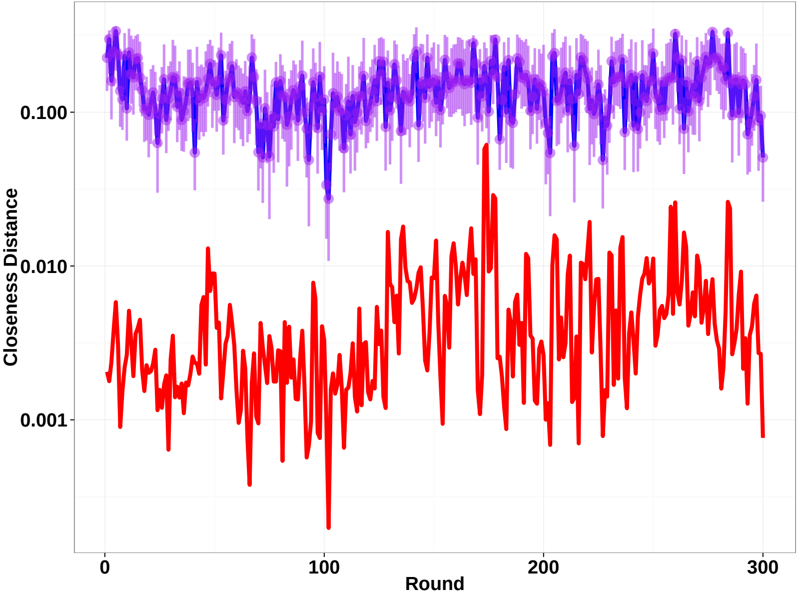
Dynamic Topologies/Results



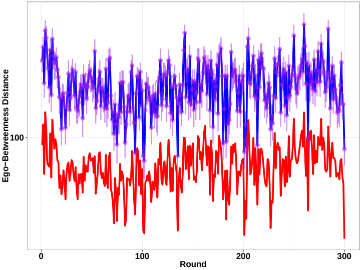
Dynamic Topologies/Results

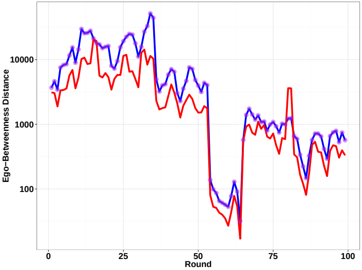


Dynamic Topologies/Results



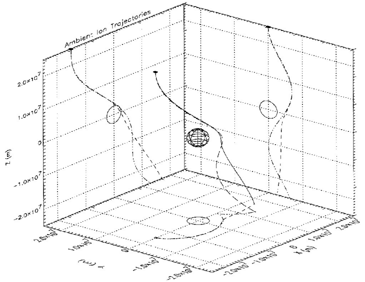
Dynamic Topologies/Results

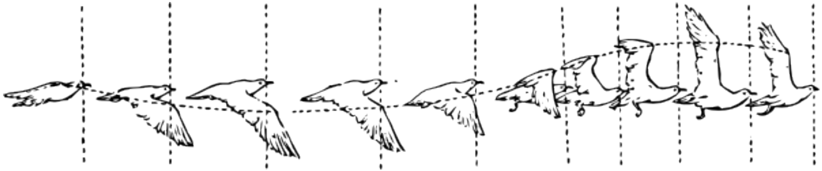




Conclusion

- GED fails to capture interesting aspects of the dynamism
- Moreover, it is overly pessimistic
- We introduce centrality distances
- Basically captures the change of roles in the network
- For each role, a centrality, and therefore a distance





Future Works

- Many, interpolation, extrapolation, *real* link prediction
- Distances among different graphs
- Measurement errors
- Find algorithms that leverage on such stability properties
- Greedy routing & GoG exploration ?