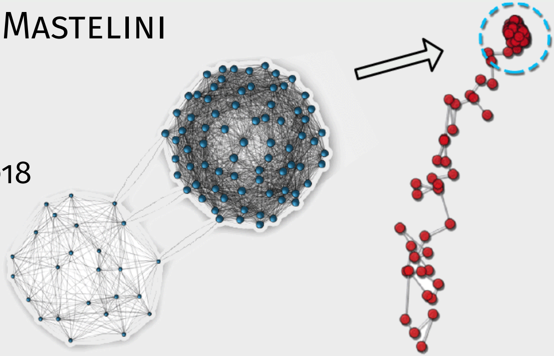
Reducing Snapshots to Points:

A Visual Analytics Approach to Dynamic Network Exploration

Saulo Martiello Mastelini

ICMC – USP

15 de Outubro de 2018



Outline

Introduction

Title:

- *Reducing Snapshots to Points: A Visual Analytics Approach to Dynamic Network Exploration*

Authors:

- Stef van den Elzen
- Danny Holten
- Jorik Blaas
- Jarke J. van Wijk

Year:

- 2016

Journal:

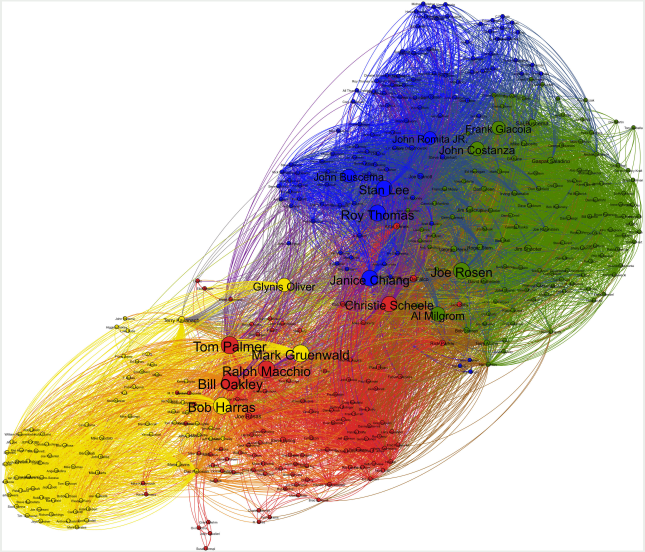
- IEEE Transactions on Visualization and Computer Graphics

Citation count:

- 56
- Graph and dynamic network visualization
- Time varying visualization

- *Networks* are ubiquitous
  - Represent relations between objects
- Often networks are large and dynamic
  - Change over time
  - Evolve!
- Examples:
  - (Tele-) communication networks
  - Social networks
  - Financial Networks
  - Transport network

Marvel Comics' Avengers artists collaboration (1963-1996)



Available in: `http://allthingsgraphed.com/public/images/marvel/avengers.svg`

Challenges

- Understanding the evolution of dynamic networks is challenging
- Discovery of states that characterize them over time:

States:

- Stable
- Recurring
- Outliers
- Transitions

Changes:

- Gradual, from a state to another
- Alternated between multiple states
- Not stable at all

Twofold challenge:

1. How to visualize, interact with, and analyze a large static network for one point in time (*snapshot*)?
2. How to visualize, interact with, and analyze many of these snapshots?

Contributions

Proposal: reducing network snapshots to points in 2D spaces

- Visual identification of network states
- Identification of transitions between states
- Analysis of network evolution in general

- Visualization tool prototype as concept proof

Related Work

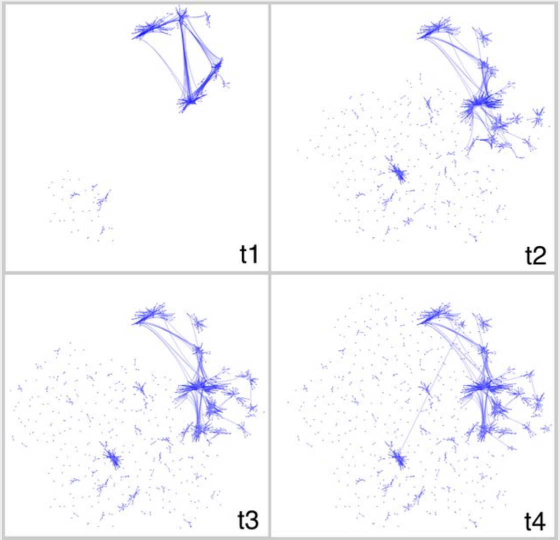
Traditional visualization approaches

1. Mapping *time-to-time*
    - Animation
2. Mapping *time-to-space*
    - Small-multiples

Animation

Available in: `https://www.youtube.com/watch?v=HP63AO9B1YM`

Small-multiples



Available in: `https://arxiv.org/pdf/1409.5034.pdf`

Current approaches

- Problems with these approaches:

Animations

High cognitive load:

- Difficulty to track changes over time
- How to visually encode these changes

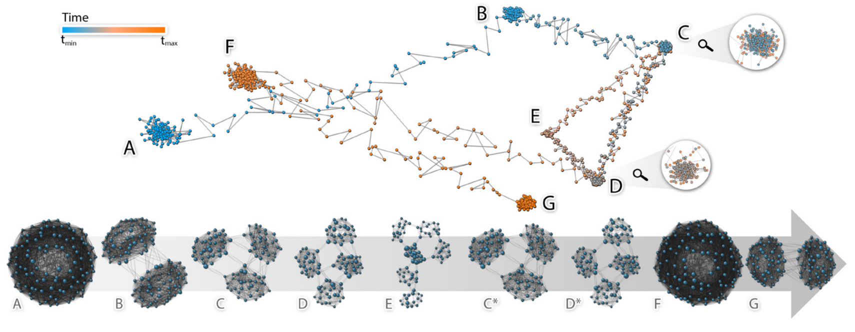
Small-multiples

To find the optimal balance between using:

Few images Lacking of temporal detail
Too many images Difficult to interpret

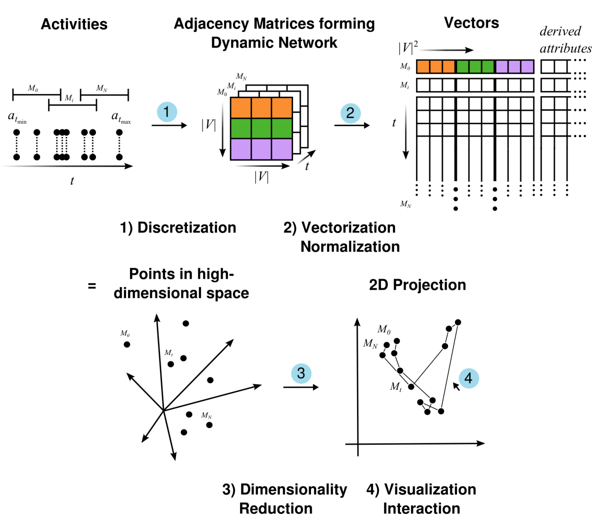
Reducing Snapshots to Points

Observe the disposition of projected network's snapshots, seeking for visible patterns



Retrieved from van den Elzen et al. [1]

Approach Overview



Adapted from van den Elzen et al. [1]

- Dynamic Network Γ as a sequence of N snapshots:

$$\Gamma = (G_1, G_2, ..., G_N)$$

- Snapshot as a directed graph:

$$G_i = (V, E_i, t_i)$$

$$\begin{aligned} V &: \text{ Node (vertex) set} \\ E_i \subseteq V \times V &: \text{ Edge (link, event) set (vertex tuples } (v_m, v_n)) \\ t_i &: \ i\text{-th timestep} \end{aligned}$$

- Complete edge set and weight function w

$$E = \bigcup_{i=1}^{N} E_i$$

$$w : i \times E \to \mathbb{R}$$

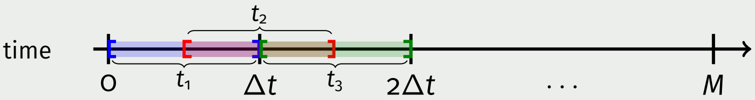
$$w(i, e), \text{ for edge } e \in E_i$$

(1) Discretization

- Discrete activities: activity log

$$A = (A_1, A_2, ..., A_j, ..., A_M) \qquad A_j \in E \times \mathbb{R}$$

- $A_j = (a_j, s_j)$ is an activity with edge (v_m, v_n) and timestamp s_j
- Time window to select snapshots with constant Δt
  - Overlapping is allowed



- Each snapshot G_i is a *adjacency matrix* M of size $|V| \times |V|$:

w_i	a	b	c	...
a	$w_{a,a}$	$w_{a,b}$	$w_{a,c}$	...
b	$w_{b,a}$	$w_{b,b}$	$w_{b,c}$	...
c	$w_{c,a}$	$w_{c,b}$	$w_{c,c}$	...
...	...	...	...	...

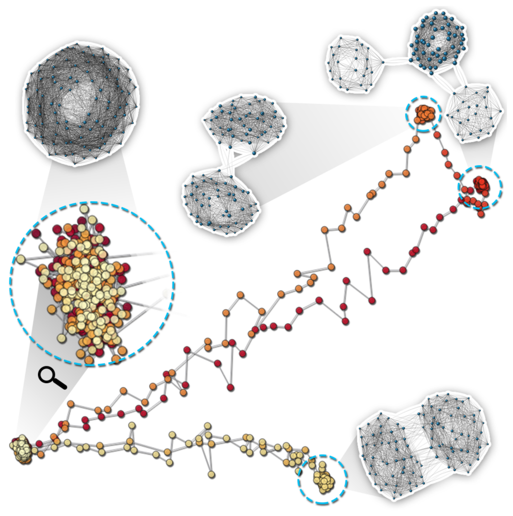
(2) Vectorization and Normalization

- Each adjacency matrix rearranged as $1 \times |V|^2$ array
- Snapshots stacked to form a $N \times |V|^2$ matrix
  - Columns represent edges, rows represent snapshots
- Different normalization strategies:
  - None
  - Binarization
  - Min-max
  - Z-score
- The choice depends on the dealt problem

(3) Dimensionality Reduction

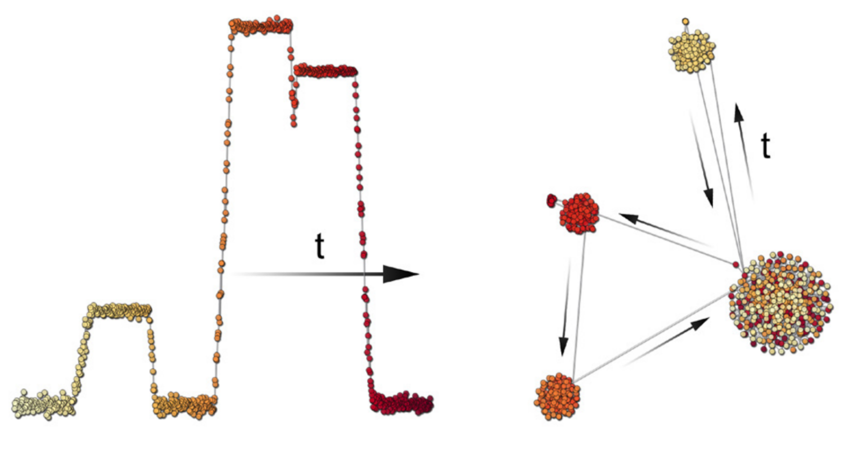
- Linear and non-linear projection techniques evaluated:
  - Principal Component Analysis (PCA)
  - t-Distributed Stochastic Neighbour Embedding (t-SNE)
- Improved versions to reduce computations:
  - *Randomized* PCA
  - *Barnes-Hut*-SNE

Example 1: Synthetic dynamic network and PCA



Retrieved from van den Elzen et al. [1]

Example 2: Time and PC1, non-linear projection



Retrieved from van den Elzen et al. [1]

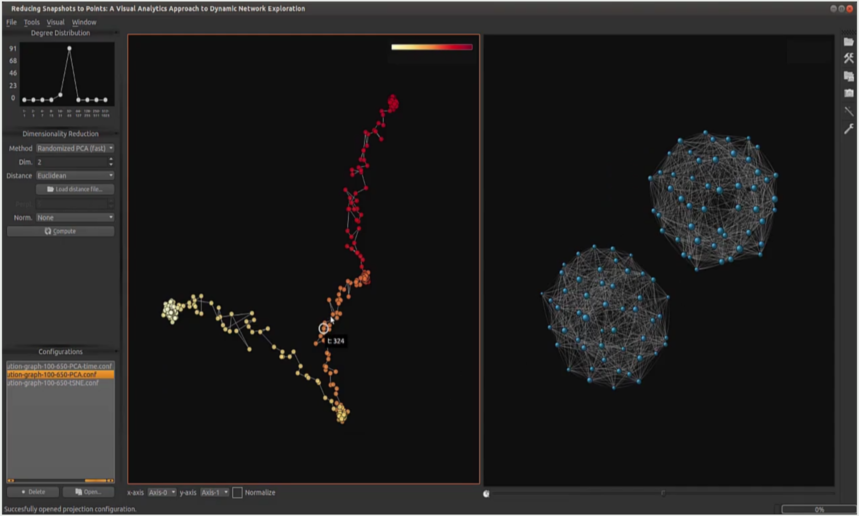
(4) Visualization Interaction

- Prototype developed as concept proof
- Two linked and juxtaposed views:

  Projection View: Each snapshot as a dot
    - Close dots represent similar network states
    - Sequential dots connected with lines
    - User defined colormap (e.g., global time, hour of the day)

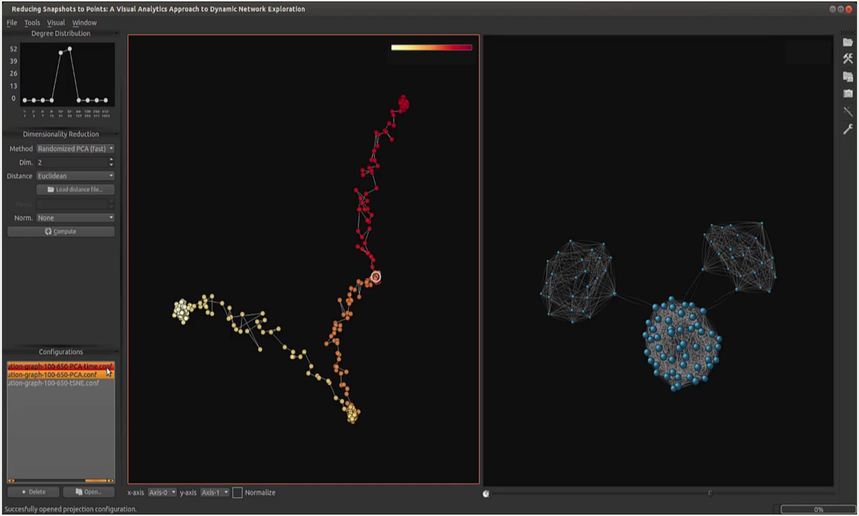
  Network View: visualization of a selected snapshot using a node-like diagram

Example 3: Projection View and Network View



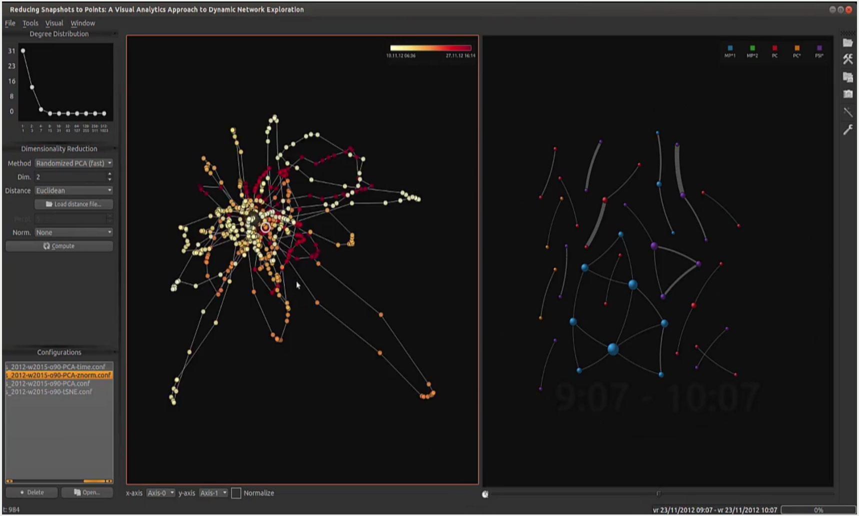
Supplementary material of van den Elzen et al. [1]

Example 4: Different projections



Supplementary material of van den Elzen et al. [1]

EXAMPLE 5: DEALING WITH DATA CLUTTERING AND ANIMATIONS IN THE NETWORK VIEW



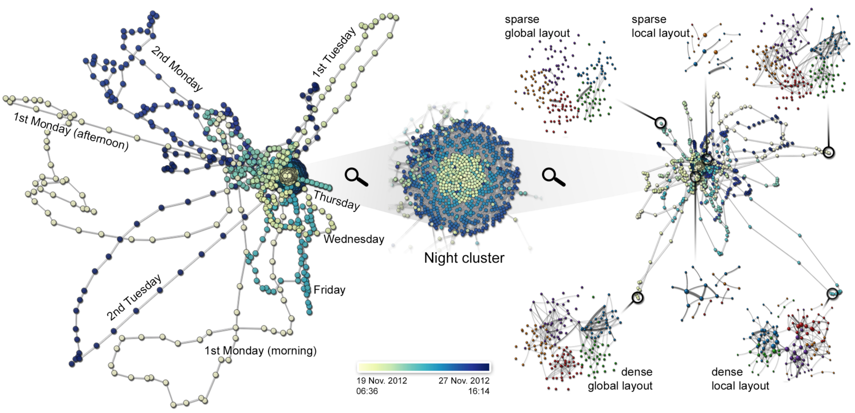
Supplementary material of van den Elzen et al. [1]

Use cases

High-school Contact Patterns

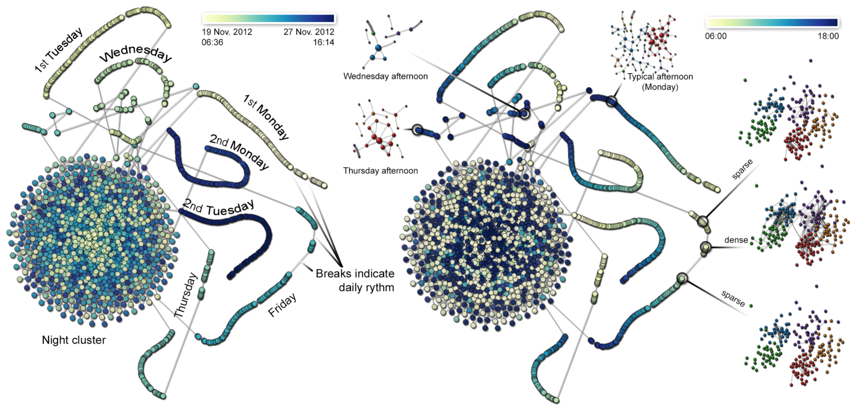
- *SocioPatterns* initiative
- Face-to-face contact between persons (wearable sensors)
  - How infectious diseases spread within a population
  - High-school students
  - 7 school days (Monday to Tuesday, in a week of 2012)
  - 180 nodes (students from 5 classes)
  - 45,047 contacts (edges)
  - 10,104 unique edges
  - Time window of 60 minutes

No normalization (left) and z-score (right)



Retrieved from van den Elzen et al. [1]

Global time variation (left) and Hour of the day (right)



Retrieved from van den Elzen et al. [1]

Discussion

Discussion

- Different tools and parameters for different problems
- Two dimensional projections:
  - Linear projections show transitions
  - Non-linear projections omit transitions but are better in separating the snapshots
- Sparse vs. Dense networks
  - Distance from the center

Conclusion

Conclusion

- Simple but yet powerful approach
- Flexibility is an advantage
- Known tools assembled in a novel approach
- Useful insights about dynamic network evolution

Text structure and quality: well written but sometimes repetitive/redundant

Software availability: prototype tool was not made available by the authors

Thanks for your time!

Any questions?

References

Stef van den Elzen, Danny Holten, Jorik Blaas, and Jarke J van Wijk.
Reducing snapshots to points: A visual analytics approach to dynamic network exploration.
IEEE transactions on visualization and computer graphics, 22(1):1–10, 2016.

Questions about the paper

1. What does the agglomeration of points with mixed colors mean?
2. What type of projection achieves the best data separation? Linear or non-linear?
3. Why is z-score the indicated normalization method to be used with PCA?