Locating Patient Zero In a Network

Pablo Curiel

CSCI - 651

December 8, 2022

Outline

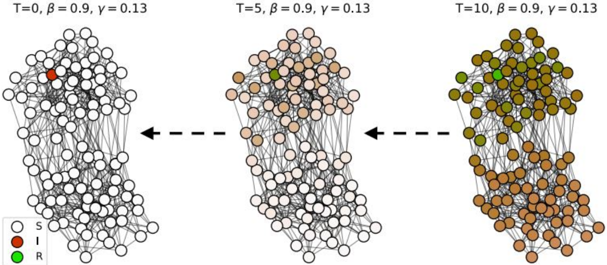
- Review of Spread in Networks
- Overview of Patient Zero (P-0) Problem
- Methods of Locating P-0
- Experimental Methods of Locating P-0
- Future Work

Review of Spread in Networks

- Applications include spread of: **disease**, rumors/fake news, malware attacks
- Classical models of spread (SI, SIS, SIR, SEIR, etc.)
  - Often represented with system of ODEs
  - Assume homogeneous mixing
- Analogs of classical models can be applied to networks
  - Hubs are super-spreaders
  - Use discrete time
- Basic reproductive number $R_0 = \frac{\beta \langle k \rangle}{\mu}$
  - $\beta =$ rate of infection, $\mu =$ rate of recovery/death, $\langle k \rangle =$ avg. degree of network
  - Number of secondary infections produced by a single infected individual in fully susceptible population
  - $R_0 > 1 \implies$ disease spread, $R_0 < 1 \implies$ disease dies out
  - COVID-19 $R_0 \approx 1.5 - 4$

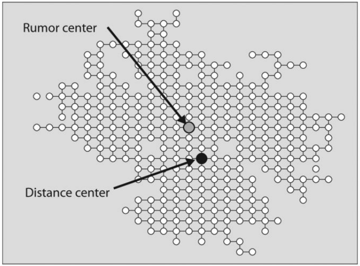
Overview of P-0 Problem

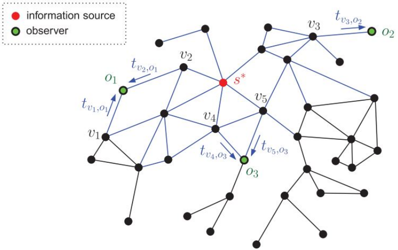
- Critical real-world application of spread in networks
  - Contact tracing
  - Allocating immunization resources
  - Mitigating spread of rumors/fake news
  - Defending against malware attacks
  - Preventing rolling blackouts when key electrical grids fail
- Somewhat newer research area (2011-Present)
- Requires learning the backwards-time dynamics of a spread
  - Often difficult and computationally expensive



Methods of Locating P-0

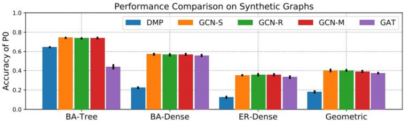
- Shah and Zaman (2011)
  - Focus on inferring source of a rumor spread using SI model
  - Define a max. likelihood estimator called ”rumor centrality”
  - Accuracy $\approx 0 - 16\%$ (depending on network structure)
- Pinto, Thiran, Vetterli (2012)
  - Make use of ”observer” nodes within network
  - Observer nodes track:
    1. Time of infection
    2. Node that infected them
  - Difficulties include observer placement/density within network
  - Complexity: $O(N)$ for arbitrary trees, $O(N^3)$ for arbitrary graphs





Methods of Locating P-0 (cont.)

- Lokhov et al. (2014)
  - Uses dynamic message passing (DMP) to infer source
  - Proven to be exact on trees
  - Complexity: $O(tN^2\langle k \rangle)$, $t =$ time-steps
- **Shah et al. (2020)**
  - Use GNNs to identify P-0 with increased efficiency/accuracy
  - Model agnostic (does not require spread model parameters)
  - Complexity: $O(N^2 logN)$
  - Accuracy $\geq\approx$ 20% (depending on graph structure and time)



Inference times

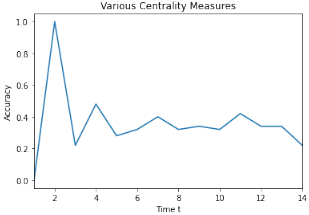
Dataset	DMP	GCN	GAT
BA-Tree	14.40 hr	3.89s	**3.18s**
BA-Dense	77.04 hr	**4.91s**	8.19s
ER-Dense	71.77 hr	**4.93s**	9.66s
Geometric	70.35 hr	**5.34s**	10.87s

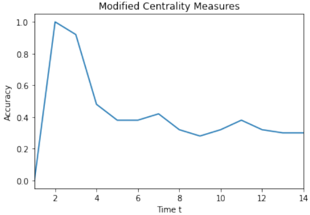
Experimental Methods of Locating P-0

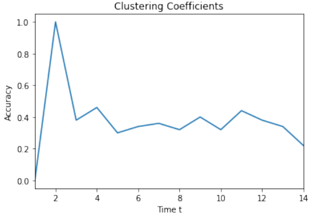
- Attempted to infer source node based on:
  - Various centrality measures (whole graph)
  - Various centrality measures (recovered subgraph)
  - Clustering coefficients
  - Centrality measures + clustering coefficients
  - Random guessing
- Inferred source was max of each measure
- Experiment parameters:
  - Time simulated = 15
  - Number of nodes N = 1000
  - Iterations per simulation = 100
  - $R_0 = 1.5$
- Graph types:
  - Barabási-Albert scale-free ($m = 3$)
  - Erdős-Rényi random ($p = 0.2$)
  - Twitter Russian trolls network ($n = 1245$, $|E| = 2974$)

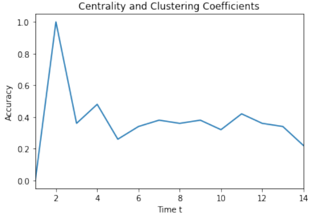
Experimental Methods (cont.)

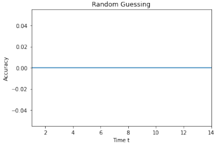
- Barabási-Albert scale-free graph results:





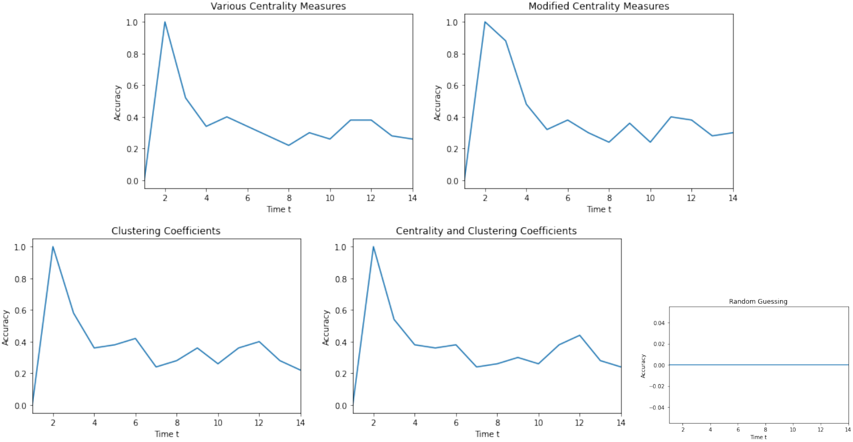




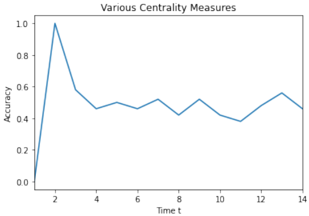


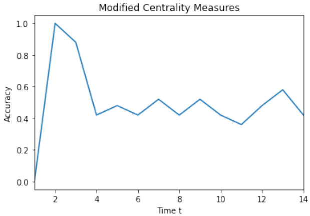
Experimental Methods (cont.)

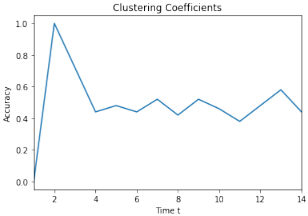
- Erdős-Rényi random graph results:

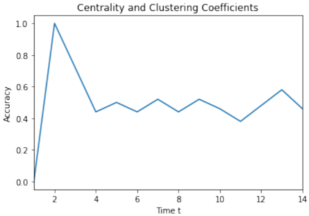


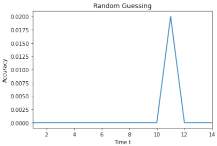
- Twitter Russian trolls graph results:











Future Work

- Understand/implement a variation of GNN-based method
  - Classify nodes based on probability of being P-0
- Expand on methods described in 2020 paper
- Focus is on GCN architecture (Kipf and Welling 2017)
  - Uses spectral convolution (Fourier Transform applied to graphs)
- Current ideas for architecture:
  - Node features: one-hot encoded node states
    - Ex: $x_f =$[1 0 0] $\implies x \in S$
  - Two node classes: P-0 ($x_c = 1$), not P-0 ($x_c = 0$)

References and Acknowledgements

- J.D. Murray. Mathematical Biology I: An Introduction. Interdisciplinary Applied Mathematics Springer, vol. 17, 2002.
- Albert-László Barabási. Network science. Cambridge University Press, 2016.
- Devavrat Shah, Tauhid Zaman. Rumors in a Network: Who's the Culprit? IEEE Transcactions on Information Theory, vol. 57, no. 8, August 2011. https://doi.org/10.48550/arxiv.0909.4370
- Pedro C. Pinto, Patrick Thiran, Martin Vetterli. Locating the Source of Diffusion in Large-Scale Networks. Phys. Rev. Lett., vol. 109, no. 6, August 2012. 10.1103/PhysRevLett.109.068702
- Andrey Y. Lokhov, Marc Mézard, Hiroki Ohta, Lenka Zdeborová. Inferring the Origin of an Epidemic with a Dynamic Message-Passing Algorithm. Phys. Rev. E, vol. 90, no. 1, July 2014. 10.1103/PhysRevE.90.012801
- Chintan Shah, Nima Dehmamy, Nicola Perra, Matteo Chinazzi, Albert-László Barabási, Alessandro Vespignani, Rose Yu. Finding Patient Zero: Learning Contagion Source with Graph Neural Networks. June 2020. https://doi.org/10.48550/arXiv.2006.11913
- Thomas N. Kipf, Max Welling. Semi-supervised Classification with Graph Convolutional Networks. ICLR 2017. https://doi.org/10.48550/arXiv.1609.02907