

Inferring the origin of an epidemic with a dynamic message-passing algorithm

HARSH GUPTA

(Based on the original work done by Andrey Y. Lokhov, Marc Mézard,
Hiroki Ohta, and Lenka Zdeborová)

Paper

- Andrey Y. Lokhov, Marc Mézard, Hiroki Ohta, and Lenka Zdeborová, [*“Inferring the origin of an epidemic with a dynamic message-passing algorithm”*](#), Physical Review E 90, 2014 – APS.
- Also available on arXiv.

Outline

- Introduction
- Infection spreading mechanism/model
 - Susceptible-infected-recovered (SIR) model
- DMP algorithm for inference
- Dynamic message-passing (DMP) equations
- Performance comparison via simulations
 - Comparing with Jordan and Distance centrality based algorithms

Introduction

- Understanding and controlling the spread of epidemics on networks of contacts is an important and significant task.
- It has far-reaching applications in mitigating the results of epidemics caused by infectious diseases, computer viruses, rumor spreading in social media, and others.
- We aim to address the specific problem of estimation of the origin of the epidemic outbreak, i.e., the so-called patient zero or infection source.

Introduction

- Information about the origin of the epidemic could be extremely useful to reduce or prevent future outbreaks (for e.g., one may fix a contaminated water body in the case of a cholera epidemic).
- **Challenge:** The only information available to us is a contact network and a snapshot of epidemic spread at a certain time. We aim to determine the infection source using this information.
- In order to define the problem, we need to know the mechanism with which the epidemic spreads. We assume the spreading mechanism follows the popular SIR model used widely in literature.

SIR model

- Susceptible-infected-recovered (SIR) model is one of the most popular and studied epidemiological models along with the susceptible-infected-susceptible (SIS) model.
- SIR model is defined as follows:
 - Let $G \equiv (V, E)$ be a connected undirected graph containing N nodes defined by the set of vertices V and the set of edges E .
 - Each node $i \in V$ at discrete time t can be in one of the three states (state variable - $q_i(t)$): susceptible $\rightarrow q_i(t) = S$, infected $\rightarrow q_i(t) = I$, or recovered $\rightarrow q_i(t) = R$.
 - The epidemic process on a graph can be interpreted as the propagation of infection signals from infected to susceptible nodes.
 - The infection signal $d^{i \rightarrow j}(t)$ is defined as a random variable which is equal to one with probability $\delta_{q_i(t-1), I} \lambda_{ij}$, and equal to zero otherwise. Here λ_{ij} measures the efficiency of spread from the node i to node j ,

SIR model

- At each time step, an infected node i will recover with probability μ_i , and a susceptible node i will become infected with probability $1 - \prod_{\{k \in \partial i\}} (1 - \lambda_{ki} \delta_{q_k(t), I})$, where ∂i is the set of neighbors of node i .
- The recovered nodes never change their state.
- In the image on the right is an example of a single instance of the inference problem generated using the SIR model.

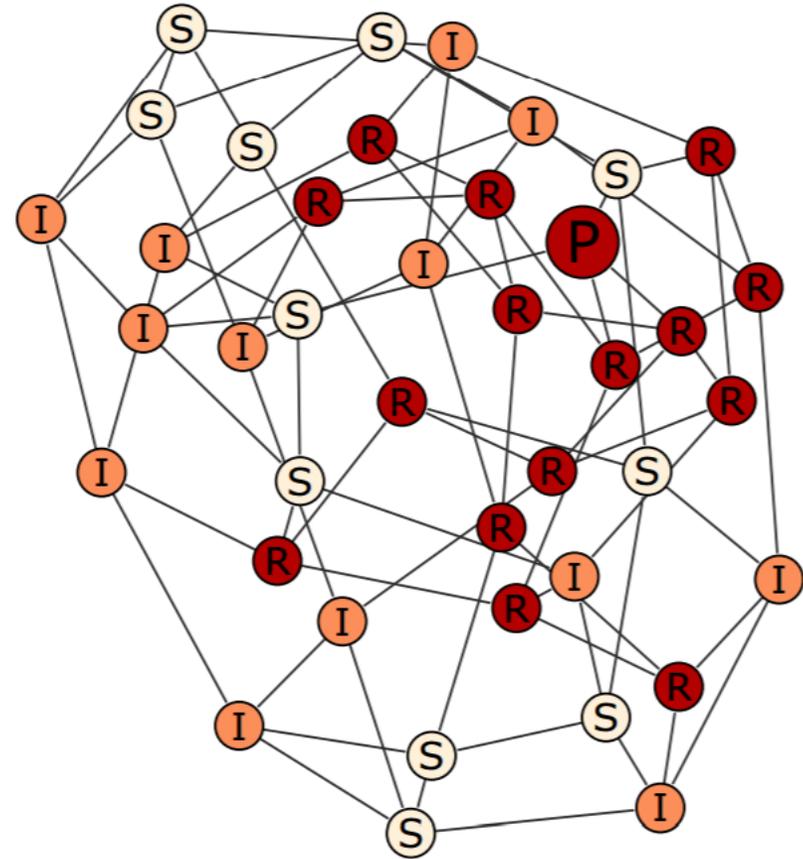


FIGURE 1

Dynamic message-passing algorithm

- The core idea behind the DMP algorithm is to infer epidemic origins using time-dependent marginal probabilities for each node in the graph. These probabilities are computed using dynamic message-passing equations.
- Let $P_S^i(t)$, $P_I^i(t)$, and $P_R^i(t)$ be the marginal probabilities that $q_i(t) = S$, $q_i(t) = I$, and $q_i(t) = R$ respectively. Let O be the snapshot of the contact network we have.
- We use DMP message-passing equations to compute the above marginal probabilities for any initial condition and time from which the snapshot might have been obtained.

Dynamic message-passing algorithm

- From the marginals, we know the probabilities $P_S^j(t, i_0)$ (respectively, $P_I^j(t, i_0)$ and $P_R^j(t, i_0)$) that a node j is in each of the states S, I, R at a given time t , for a given patient zero i_0 .
- For now, let us assume that we know the time t_0 at which the snapshot was obtained. Using Bayes rule, $P(i|O, t_0) \propto P(O|i, t_0)$.
- Intuitively, we compute the above posterior probability for all nodes i , and choose the node which maximizes this probability as patient zero.

Dynamic message-passing algorithm

- There is no tractable way to compute exactly the joint probability of the observed snapshot O . We approximate it using a mean-field type approach:

$$P(O|i, t_0) \approx \prod_{k \in O, q_k(t_0)=S} P_S^k(t_0, i) \prod_{l \in O, q_l(t_0)=I} P_I^l(t_0, i) \prod_{m \in O, q_m(t_0)=R} P_R^m(t_0, i)$$

- If the value of t_0 is not known, we estimate it by choosing the value that maximizes the partition function $Z(t) = \sum_i P(O|i, t)$.

DMP equations- How to compute marginals?

- Since the aforementioned marginals sum to one, therefore:

$$P_I^i(t + 1) = 1 - P_S^i(t + 1) - P_R^i(t + 1)$$

- As the recovery process from the state I to state R is independent of neighbors, for the SIR model we have:

$$P_R^i(t + 1) = P_R^i(t) + \mu_i P_I^i(t)$$

- We need one more update equation, i.e., for $P_S^i(t + 1)$ in order to complete the above set of equations. We will define 3 separate messages under different auxiliary dynamics to obtain an update rule for $P_S^i(t + 1)$.

Dynamic message-passing equations

- We consider two kinds of auxiliary dynamics:
 - D_j where node j receives infection signals, but ignores them and thus is fixed to the S state at all times.
 - D_{ij} where neighboring nodes i and j receive infection signals, but ignore them and are fixed to the S state at all times.
- We note that the auxiliary dynamics D_j is identical to the original dynamics for all times such that $q_j(t) = S$ (similar analogy for D_{ij}).
- Also note that in dynamics D_j , since the infection cannot propagate through node j , different graph branches rooted at node j become independent if the underlying graph is a tree.

Dynamic message-passing equations

- To obtain a closed system of marginal probability update equations, we will write the update rules for three different kinds of messages.
- The message $\theta^{k \rightarrow i}(t)$ is the probability that the infection signal has not been passed from node k to node i up to time t in the dynamics D_i :

$$\theta^{k \rightarrow i}(t) = \text{Prob}^{D_i} \left[\sum_{t'=0}^t d^{k \rightarrow i}(t') = 0 \right]$$

- The message $\phi^{k \rightarrow i}(t)$ is the probability that the infection signal has not been passed from node k to node i up to time t in the dynamics D_i and that node k is in the state I at time t :

$$\phi^{k \rightarrow i}(t) = \text{Prob}^{D_i} \left[\sum_{t'=0}^t d^{k \rightarrow i}(t') = 0, q_k(t) = I \right]$$

Dynamic message-passing equations

- Finally, the message $P_S^{k \rightarrow i}(t)$ is the probability that node k is in the state S at time t in the dynamics D_i :

$$P_S^{k \rightarrow i}(t) = \text{Prob}^{D_i}[q_k(t) = S]$$

- From the definition of the messages, we have:

$$P_S^i(t) = P_S^i(0) \prod_{k \in \partial i} \theta^{k \rightarrow i}(t + 1)$$

where ∂i means the set of neighbors of i .

- It can be shown easily that:

$$P_S^{i \rightarrow j}(t) = P_S^i(0) \prod_{k \in \partial i \setminus j} \theta^{k \rightarrow i}(t + 1)$$

Dynamic message-passing equations

- For the remaining two messages, the update rules are as follows:

$$\theta^{k \rightarrow i}(t + 1) - \theta^{k \rightarrow i}(t) = -\lambda_{ki} \phi^{k \rightarrow i}(t), \quad \theta^{i \rightarrow j}(0) = 1.$$

$$\phi^{k \rightarrow i}(t + 1) = (1 - \lambda_{ki})(1 - \mu_k) \phi^{k \rightarrow i}(t) - [P_S^{k \rightarrow i}(t + 1) - P_S^{k \rightarrow i}(t)],$$
$$\phi^{k \rightarrow i}(0) = \delta_{q_i(0), I}.$$

- Combining the message-passing update equations, we get an iterative update rule for $P_S^i(t + 1)$. This completes the set of equations required to compute all the marginal probabilities.

Experimental Results

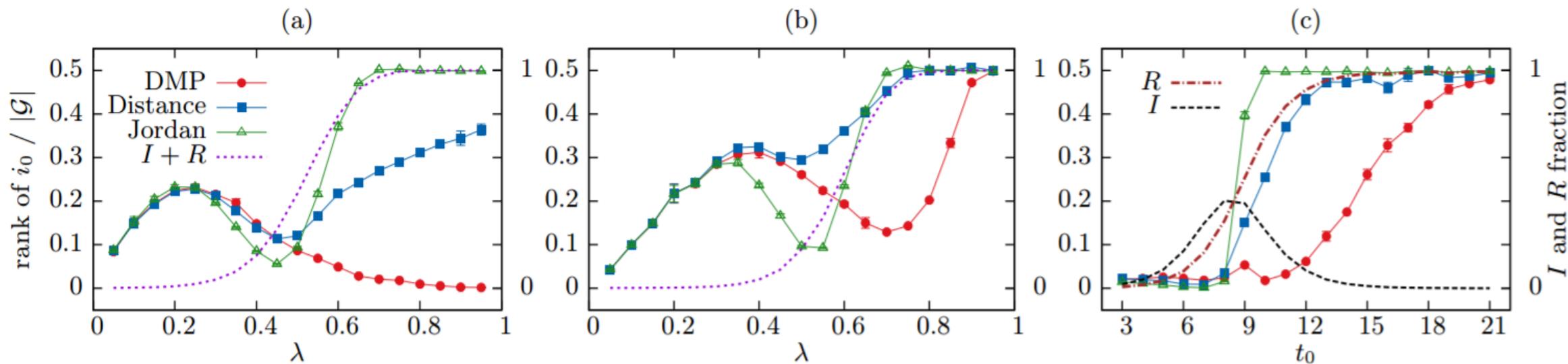


FIG. 3. (color online) Average rank of the true epidemic origin on random regular graphs of size $N = 1000$ with degree $c = 4$. Each data point is averaged over 1000 instances. The plots (a) and (b) represent the dependences of the average rank on the infection rate λ , for the snapshot time $t_0 = 10$ and recovery probability μ : (a) $\mu = 0.5$ and (b) $\mu = 1$. In this figure t_0 is inferred by the algorithm. The DMP estimator (red circles) is compared to the Jordan centrality (green triangles) and the distance centrality (blue squares) estimators. The dotted line shows the average fraction of nodes that were infected or recovered in the snapshot, $|\mathcal{G}|/N$, we use this number to normalize the ranks of the epidemic origin. Plot (c) shows the dependence on the snapshot time t_0 for $\lambda = 0.7$ and $\mu = 0.5$. The dashed line is the average fraction of nodes that were infected and the dash-dotted line is the average fraction of nodes that were recovered in the snapshot; both are normalized to N .

Experimental Results

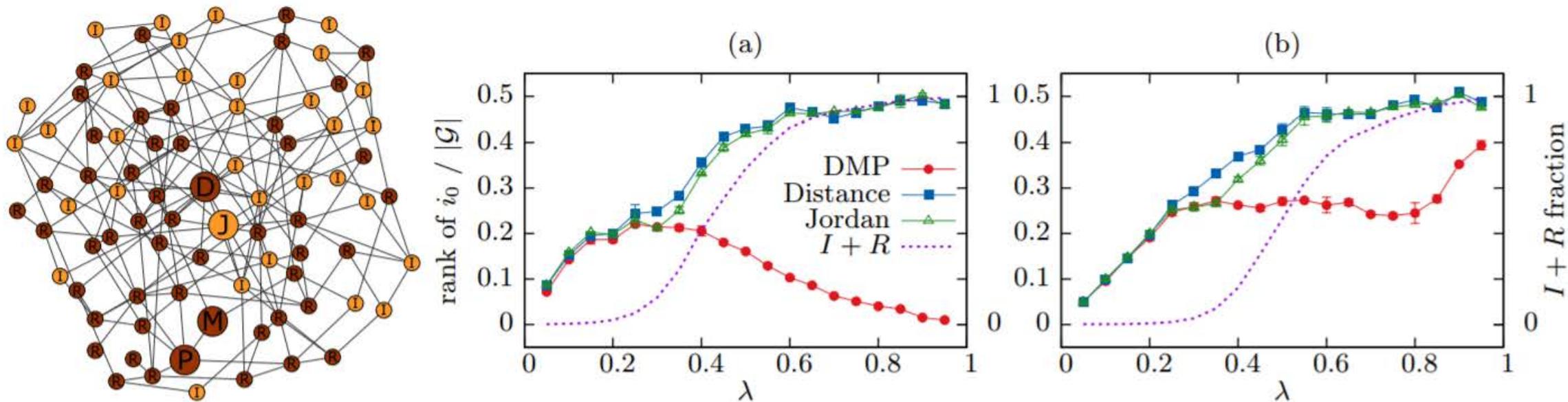


FIG. 4. (color online) Left: An instance of inference problem on the Erdős-Rényi graph with average degree $\langle c \rangle = 4$ and $N = 84$. The epidemic is generated for $\lambda = 0.7$ and $\mu = 0.5$. In this example, only infected (light) and recovered (dark) nodes are present in the snapshot at time $t_0 = 5$. The true patient zero is labeled by P , the best-ranked nodes for DMP (M), Jordan (J) and distance (D) centralities are at distances 1, 2 and 3 from P , correspondingly. Right: Average rank of the true epidemic origin on Erdős-Rényi graphs of size $N \simeq 1000$ with average degree $\langle c \rangle = 4$. Each data point is averaged over 1000 instances. The snapshot time t_0 (assumed to be known) and recovery probability μ are: (a) $t_0 = 10$, $\mu = 0.5$ and (b) $t_0 = 10$, $\mu = 1$. The dotted line shows the average fraction of nodes that were infected or recovered in the snapshot, $|\mathcal{G}|/N$, we use this number to normalize the ranks of the epidemic origin.

Thank you