Node Immunization with Non-backtracking Eigenvalues and X-Centrality

Northeastern University Network Science Institute

Leo Torres* (leo@leotrs.com), Kevin S. Chan, Hanghang Tong, and Tina Eliassi-Rad

arxiv.org/abs/2002.12309

Abstract

- Spreading processes are characterized by a reproductive number R_0. If R_0 is larger than a threshold θ, the process becomes an **epidemic**.
- It is known that $\theta \approx 1/\lambda$, where λ is the largest eigenvalue of the **non-backtracking matrix**.
- **Node immunization** removes nodes from the network to minimize λ (i.e. maximize θ).
- Let λ_c be the eigenvalue after immunizing node c. Then, $\lambda - \lambda_c$ **is the eigen-gap of** c.
- The best targets for immunization are nodes with **large eigen-gaps**. However, computing the eigen-gap is inefficient.
- **In this work**, we describe ways of identifying **which nodes have the largest eigen-gaps**.

X-Centrality

Let z be a vector indexed by the oriented edges of the graph, and v be the principal eigenvector of B, then

- If $z = v$, $\sum_j z_{j \to i}$ equals the **non-backtracking centrality** of i.
- If $z = \mathbf{1}$, $\sum_j z_{j \to i}$ equals the **degree** of i.
- For arbitrary z, the **X-centrality** of c is

$$z^T P X z = \left(\sum_i a_{ic} \left(\sum_j z_{j \to i}\right)\right)^2 - \sum_i a_{ic} \left(\sum_j z_{j \to i}\right)^2$$

Intuitively, the X-centrality measures the **diversity** of the values of the expression $\sum_j z_{j \to i}$ among neighbors of c.

Algorithm: immunization with X-degree

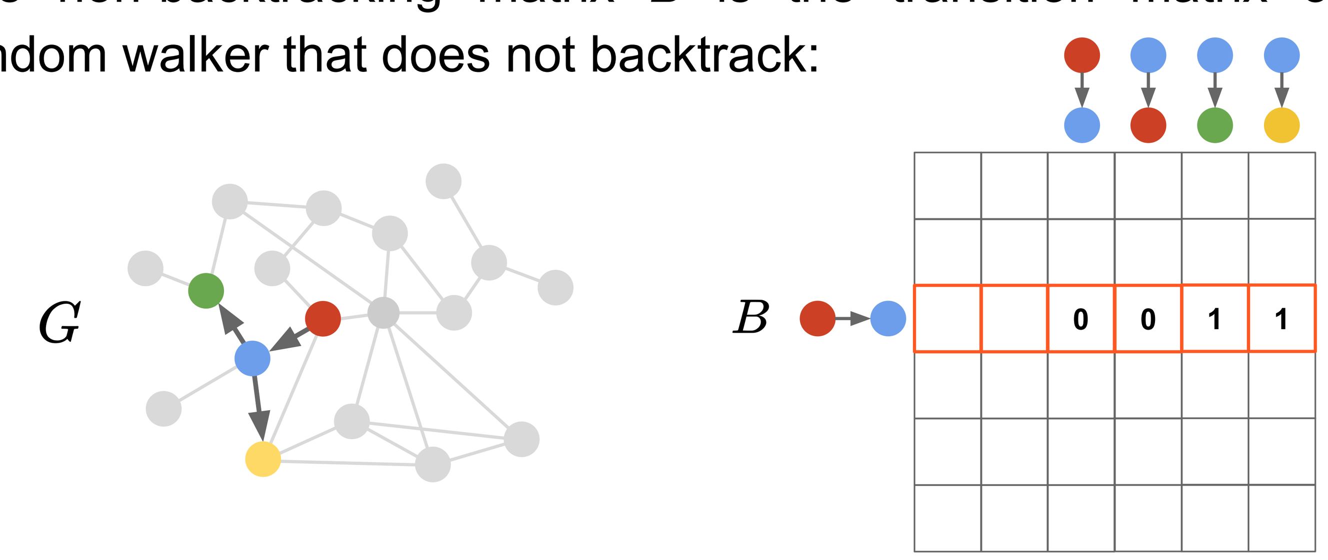
Input: graph G, integer p
Output: removed, an ordered list of nodes

```
removed ← ∅
XDeg[i] ← XDegree(G, i) for each node i
while length(removed) < p do
    node ← max_i XDeg[i]
    foreach i in G.neighbors[node] do
        G.neighbors[i].remove(node)
    foreach i in G.neighbors[node] do
        foreach j in G.neighbors[i] do
            XDeg[j] ← XDegree(G, j)
    G.neighbors[node] ← ∅
    removed.append(node)
return removed
```

This algorithm can be implemented using one of two data structures: **an indexed priority queue (IPQ), or a hash table (a.k.a. dictionary, Map)**. Each version is more efficient on **different types of networks.**

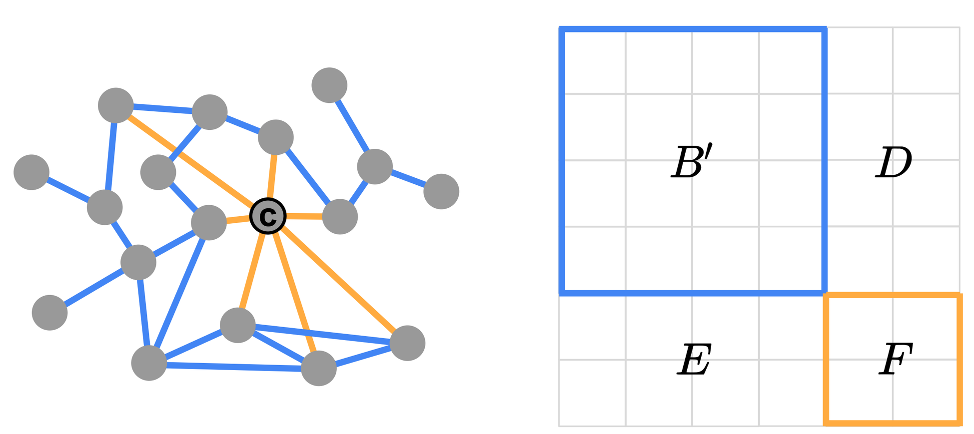
Perturbation of non-backtracking eigenvalues

The non-backtracking matrix B is the transition matrix of a random walker that does not backtrack:

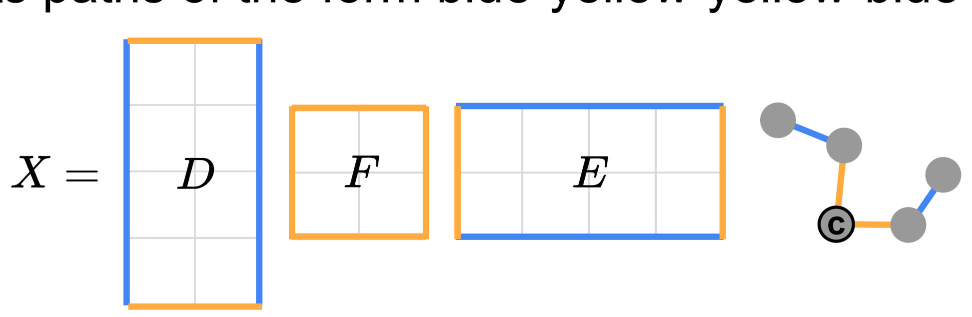


Strategy: estimate the eigen-gap of c using the following

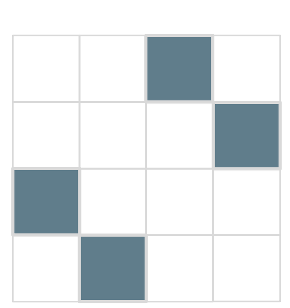
- Block form



- X counts paths of the form blue-yellow-yellow-blue



- P is a permutation matrix



Estimating the eigen-gap: theoretical results

Theorem: If v is the principal eigenvector of B', then

$$\lambda^2 (\lambda - \lambda_c) \approx v^T P X v$$

Definition: $v^T P X v$ is called the **X-non-backtracking centrality** of c.

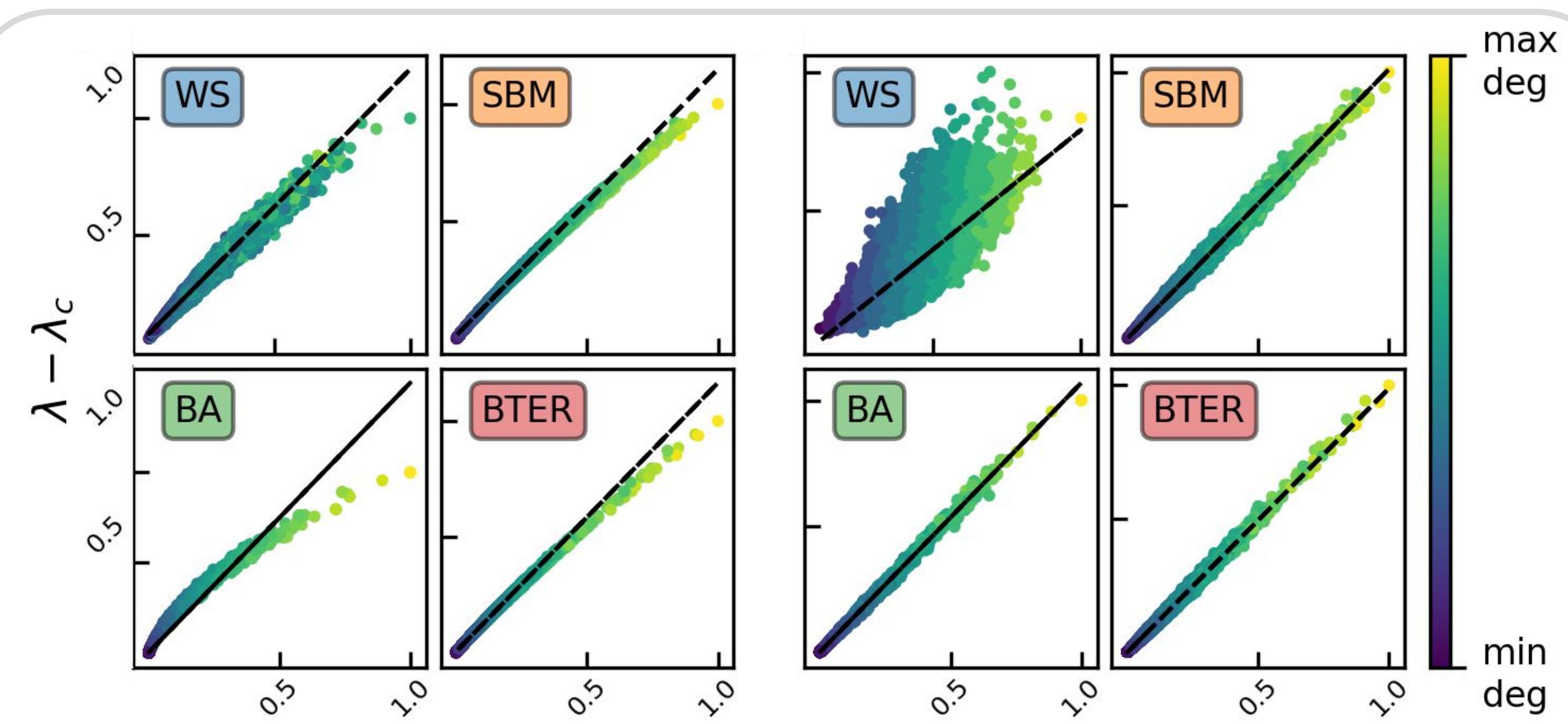
Theorem: The eigen-gap is upper bounded by

$$\lambda^2 (\lambda - \lambda_c) \leq \left(\mathbf{1}^T P X \mathbf{1}\right)(Q_c)$$

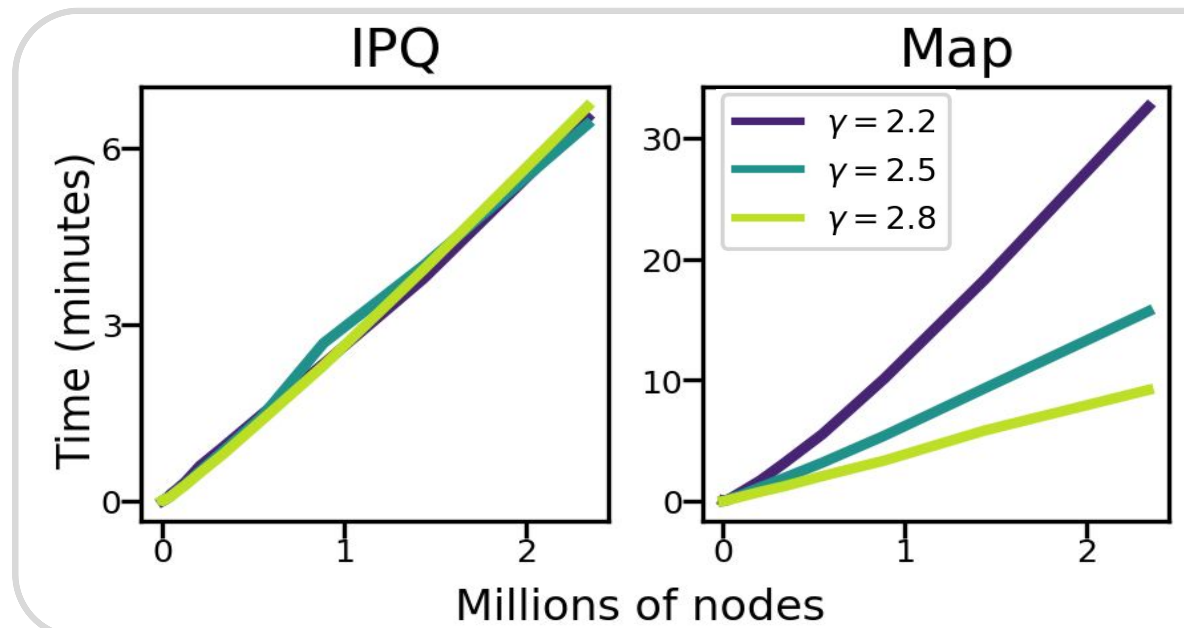
where Q_c is a constant close to 1.

Definition: $\mathbf{1}^T P X \mathbf{1}$ is called the **X-degree centrality** of c.

X-NB and X-degree can predict the eigen-gap



Linear scalability



Immunization on **random configuration model graphs** whose degree distribution is sampled from a power-law distribution with exponent γ.

X-degree is more effective than baselines

	$p = 1$ degree	CI	Xdeg	$p = 10$ degree	CI	Xdeg	$p = 100$ degree	CI	Xdeg
AS-1	0.74	0.74	**2.35**	6.70	13.51	**15.43**	71.65	**78.26**	75.92
AS-2	2.02	2.02	**4.00**	17.09	22.36	**28.17**	87.60	**89.61**	87.02
Social-Slashdot	0.95	**1.02**	**1.02**	4.63	6.06	**6.94**	23.65	28.11	**30.30**
Social-Twitter	**2.18**	**2.18**	1.98	13.21	**13.97**	13.68	41.10	42.88	**43.39**
Transport-California	0.00	0.00	**0.65**	**2.65**	0.65	**2.65**	5.09	5.09	**7.80**
Transport-Sydney	**0.00**	**0.00**	**0.00**	0.00	0.00	**6.50**	0.00	7.37	**9.49**
Web-NotreDame	**9.34**	**9.34**	**9.34**	12.10	**13.79**	**13.79**	14.37	14.37	**19.22**

Average percentage eigen-drop on real networks (larger is better) when removing $p = 1, 10,$ or 100 nodes. degree: **baseline,** CI: **Collective Influence algorithm,** Xdeg = **X-degree**

Conclusions and future work

1. The **X-centrality** framework provides an excellent way of performing **node immunization** by **estimating eigen-gaps**.
2. What other **X-centralities** can we define? Which are meaningful?
3. What about **edge** removal? **Subgraph** removal?
4. Non-backtracking eigenvalues have many applications: community detection, graph distance, etc. **How do these change after node removal?**

Supported in part by NSF IIS-1741197 and the Combat Capabilities Development Command Army Research Laboratory under Cooperative Agreement Number W911NF-13-2-0045.