# Finding regions with high amount of directed *c*-cycles in digraphs with complex-valued eigenpairs

Christine Klymko

Lawrence Livermore National Laboratory
Center for Applied Scientific Computing

Graph Exploitation Symposium
15-16 July 2015

# Highly cyclic structure in directed graphs



- Cyclic and highly-cyclic substructure has been shown to be an important feature in many types of complex networks.
    - web networks
    - protein-protein interaction networks
    - citation networks
    - economic networks
    - etc...
- 2-cyclic (directed bipartite) are the most well-known, but areas with larger cycles also occur.
- Often these regions are not “purely” cyclic, but instead are connected to (and hidden within) a larger network structure.

# Identifying highly-cyclic regions



- Identifying the existence of highly cyclic regions in complex, real-world networks is not easy.
- When edges are directed, this becomes more complicated.
- This work focuses on finding highly 3-cyclic regions, but the same methods apply to cycles of any length.

# Definitions

- directed graph (digraph): $G = (V, E)$
  $|V| = n$ vertices (nodes)
  $E = \{(i,j) | i,j \in V\}$, $|E| = m$ edges
- in-degree of vertex $i$: $d_i^{in}$ = # of edges entering node $i$
- out-degree of vertex $i$: $d_i^{out}$ = # of edges leaving node $i$
- *walk* of length $k$: sequence of vertices $i_1, i_2, \ldots, i_{k+1}$ such that $(i_, i_{\ell+1}) \in E$ for all $1 \leq \ell < k$.
- $q$-th roots of unity, for $p = 0, 1, ..., (q-1)$:

$$\theta_{p,q} = \exp\left(\frac{p}{q} 2\pi\iota\right)$$

where the complex unit $\iota$ satisfies $\iota^2 = -1$

# Definitions



- adjacency matrix of $G$:

$$A = (a_{ij}); \quad a_{ij} = \begin{cases} 1, & \text{if } (i,j) \text{ is an edge in } G, \\ 0, & \text{else.} \end{cases}$$

- out-degree matrix of $G$:

$$D = A\mathbf{1} = (d_{ij}); \quad d_{ij} = \begin{cases} d_i^{out}, & \text{if } i = j, \\ 0, & \text{else.} \end{cases}$$

- stochastic transition matrix of $G$:

$$B = D^{-1}A; \quad b_{ij} = \begin{cases} \frac{1}{d_i^{out}}, & \text{if } (i,j) \in E, \\ 0, & \text{else.} \end{cases}$$

# Purely 3-cyclic structure



- When a graph is known to be purely 3-cyclic, there are simple, walk-based methods for separating the three clusters.
- The eigenvalues of the (row) stochastic transition matrix $B$ can be used to identify whether a graph has a purely 3-cyclic structure.

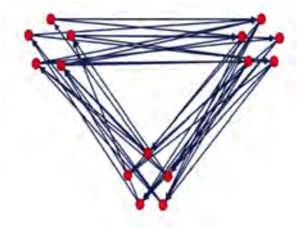


Figure : A purely 3-cyclic graph.

# Purely 3-cyclic structure

**Theorem 1.**

*Let the graph associated with $B$ be strongly connected. Then, $\theta_{1,3} \in \sigma(B)$ if and only if the graph associated with $B$ is $3k$-cyclic, for some positive integer $k$*

**Theorem 2.**

*Let $G$ be a strongly connected $3k$-cyclic graph with stochastic transition matrix $B = D^{-1}A$ that can be written in block form*

$$B = \begin{bmatrix} 0 & 0 & C_3 \\ C_1 & 0 & 0 \\ 0 & C_2 & 0 \end{bmatrix}.$$

*Further, let $\mathbf{x}$ be an eigenvector of $B$ associated with eigenvalue $\lambda = \theta_{1,3}$. Then, the entries of $\mathbf{x}$ cluster the nodes in the network according to their membership in $C_1$, $C_2$, or $C_3$.*

# Purely 3-cyclic structure

**Lemma 3.**

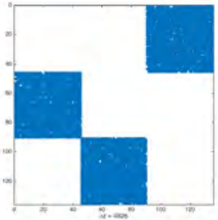
*Let $(\lambda, \mathbf{v})$ and $(\overline{\lambda}, \overline{\mathbf{v}})$ be eigenpairs of $B$ such that $Im\,\lambda \neq 0$. Let $\mathbf{v} = \mathbf{r} + \iota\,\mathbf{c}$. Consider the two-dimensional spectral coordinates $(v_i, \overline{v_i})$. There exists a 2d complex orthogonal rotation that places these coordinates in $\mathbb{R}^2$:*

$$\mathbf{x} = \frac{1}{\sqrt{2}}(\mathbf{v} + \overline{\mathbf{v}}) = \sqrt{2}\,\mathbf{r} \qquad \textit{and} \qquad \mathbf{y} = \frac{1}{\sqrt{2}\iota}(\mathbf{v} - \overline{\mathbf{v}}) = \sqrt{2}\,\mathbf{c}.$$

In the case where $\lambda = \theta_{1,3}$, this embedding separates and identifies the three groups of nodes.

# Purely 3-cyclic structure





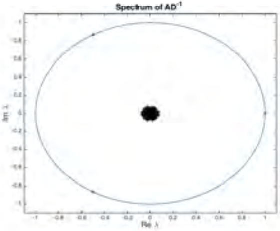



Figure : The adjacency matrix (left) and spectrum (right) of a purely 3-cyclic graph.

# Purely 3-cyclic structure



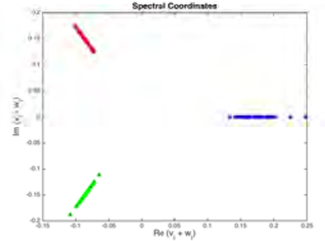



Figure : Coordinate embeddings using $\theta_{1,3}$.

# Locally 3-cyclic structure



- When even a little bit of noise is added to a purely 3-cyclic graph, walk based methods for identifying regions start to fail.
- The eigenvalues and eigenvectors of $B$ can still provide a lot of information.

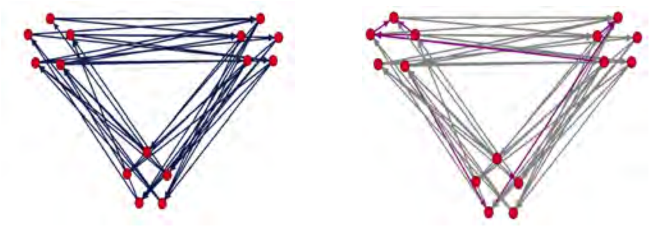


Figure : A 3-cyclic graph with some noise added.

# Locally 3-cyclic structure



**Theorem 4.**

*Let $G$ be a strongly connected graph such that the stochastic transition matrix of $G$, $B = D^{-1}A$ can be written in block form*

$$B = \begin{bmatrix} G_o & & & \\ & 0 & 0 & C_3 \\ & C_1 & 0 & 0 \\ & 0 & C_2 & 0 \end{bmatrix}$$

*where $G_o$ is strongly connected and is not 3k-cyclic for any $k$. Let $\hat{B} = D^{-1}A + M$ be a row stochastic matrix representing the original matrix plus noise. Then, there exists $\lambda \in \sigma(\hat{B})$ such that*

$$|\lambda - \theta_{1,3}| < \frac{2\sqrt{n}}{3} \left( \max_i \frac{\hat{d}_i - d_i}{\hat{d}_i} \right)$$

*where $\hat{d}_i$ is the out-degree of node $i$ in the graph with noise and $d_i$ is the out-degree of node $i$ in the noiseless graph.*

# Locally 3-cyclic structure



**Theorem 5.**

*Let $\hat{B}$ be as stated in Theorem 4. Then, there exist eigenpairs of $\hat{B}$ $(\lambda_o, \mathbf{x})$ and $(\overline{\lambda_o}, \overline{\mathbf{x}})$ which can be used to define a mapping into $\mathbb{R}^2$ such that each node $j$ is mapped into a circle of radius*

$$r = 1 + (1-\epsilon)^{2k} \left(1 - \epsilon \left(\frac{d_{max} - (1-\epsilon)^{-1}}{d_{max} - 1}\right)\right)^{2k-2} \tag{1}$$

$$-2(1-\epsilon)^{k-1} \left(1 - \epsilon \left(\frac{d_{max} - (1-\epsilon)^{-1}}{d_{max} - 1}\right)\right)^{k-1} (1 - \epsilon d_{max} + \frac{1}{2}\epsilon^2 d_{max}) \tag{2}$$

*around vectors of length 1 at angles of $\frac{\pi}{3}, \pi$ or $\frac{5\pi}{3}$ where $d_{max}$ is the maximum degree in the network and $k$ is the number of steps on the shortest path between node $i$ associated with $|x_i| \geq |x_l|$ and node $j$.*

# Locally 3-cyclic structure



In practice, Theorems 4 and 5 indicate that:

- As noise is added to a $3k$-cyclic network, there exists $\epsilon > 0$ and an eigenvalue $\lambda_o$ such that $|\lambda_o - \theta_{1,3}| < \epsilon$ .
- The eigenvector $\mathbf{x}$ associated with $\lambda_o$ is guaranteed to identify a seed node in each of the three sets.

When the $3k$-cyclic region is well separated from the larger network (few "noise" edges), $\mathbf{x}$ will often identify the three sets completely.

# Tests on stochastic block models



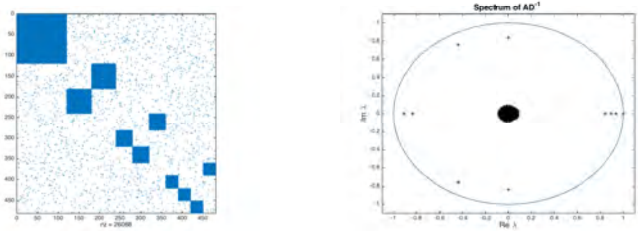



Figure : The adjacency matrix (left) and spectrum (right) of a test problem.

# Tests on stochastic block models

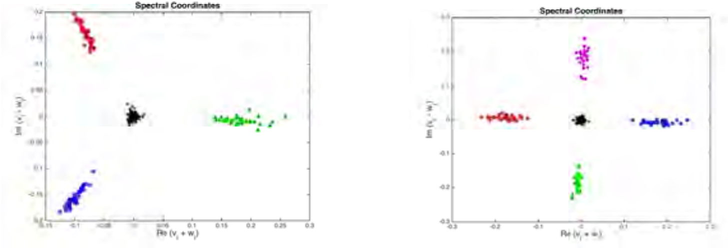



Figure : Coordinate embeddings using $\theta_{1,3}$ (left) and $\theta_{1,4}$ (right). Black nodes were unclustered in these embeddings.

# Conclusions



- The eigenvalues and eigenvectors of $B$ can be used to identify highly cyclic regions in complex networks.
- Additional tests on both generated and real world networks are needed.
- Future work includes determining methods for quickly estimating the eigenvalues/vectors of interest and extending the analysis to overlapping communities.

# (Partial) Bibliography



E. ESTRADA, *Spectral scaling and good expansion properties in complex networks* Europhys. Lett. 73, 649 (2006).

H. KIM AND J. M. KIM, *Cyclic topology in complex networks*, Phys. Rev. E 72, 036109 (2005).

M. MIDDENDORF, E. ZIV, AND C. H. WIGGINS, *Inferring network mechanisms: The Drosophila melanogaster protein interaction network* PNAS 102, 9, pp. 3192-3197 (2005).

A. VÁZQUEZ, J. G. OLIVEIRA, AND A. L. BARABÁSI, *Inhomogeneous evolution of subgraphs and cycles in complex networks* Phys. Rev. E 71, 025103(R) (2005.