

Geometry of Exponential Random Graph Models

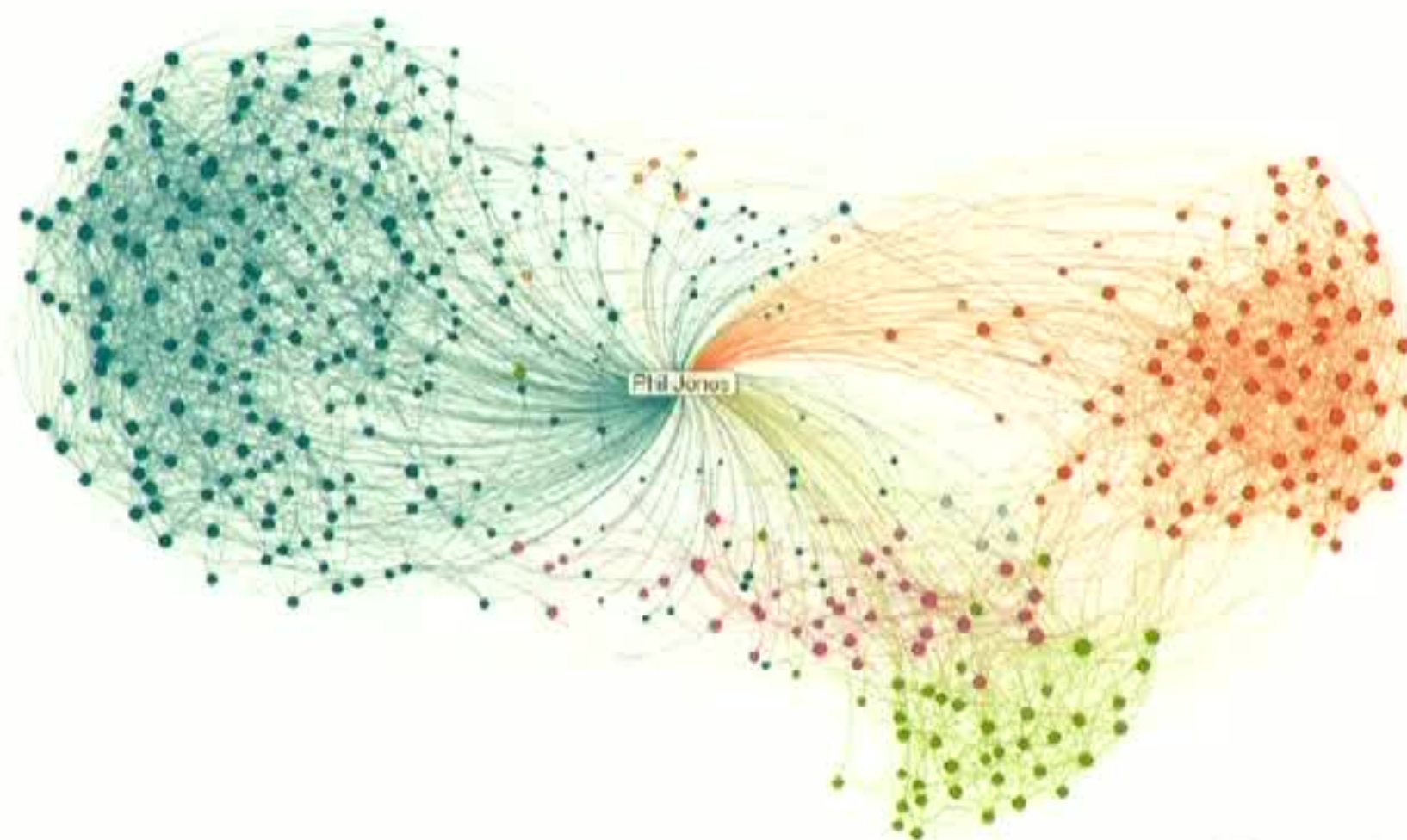
Rodolfo Garcia, Elizabeth Gross
San José State University

Ha Khanh Nguyen
The Ohio State University

Christopher O'Neill
University of California Davis

Big Statistical Challenge: *Researchers want to know how well a proposed model fits an observed network.*

- One of the most studied classes of network models are exponential random graph models, which are defined by network statistics (sufficient statistics).



Background and Notation

Adjacency Matrix: Let $G = (V, E)$ be a graph with n vertices. Let A be the adjacency matrix of the graph, i.e.

$$A_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

Network Statistics: The sufficient statistics of the edge-2path-triangle model are:

- The number of edges of G : $t_1 = \sum_{i < j} A_{ij}$.
- The number of 2-paths in G : $t_2 = \sum_{i < j} (A^2)_{ij}$.
- The number of triangles in G : $t_3 = \frac{1}{6} \text{tr} A^3$.

Background and Notation

Adjacency Matrix: Let $G = (V, E)$ be a graph with n vertices. Let A be the adjacency matrix of the graph, i.e.

$$A_{ij} = \begin{cases} x_{ij} & \text{if } i < j, \\ 0 & \text{if } i = j, \\ x_{ji} & \text{if } i > j. \end{cases} \quad (2)$$

Network Statistics: The sufficient statistics of the edge-2path-triangle model are:

- The number of edges of G : $t_1 = \sum_{i < j} A_{ij}$.
- The number of 2-paths in G : $t_2 = \sum_{i < j} (A^2)_{ij}$.
- The number of triangles in G : $t_3 = \frac{1}{6} \text{tr} A^3$.

Background and Notation

Given a graph G and network statistic $t = T(G) = (t_1, t_2, t_3)$, the reference ideal of t is:

$$\begin{aligned} I_t = \langle & \sum_{i < j} x_{ij} - t_1, \\ & \sum_{i < j < k} (x_{ij}x_{jk} + x_{ij}x_{ik} + x_{ik}x_{jk}) - t_2, \\ & \sum_{i < j < k} x_{ij}x_{jk}x_{ik} - t_3 \rangle \\ & \subseteq k[x_{ij} \mid 1 \leq i < j \leq n]. \end{aligned}$$

We will call the variety defined by I_t the reference variety:

$$V_t = V(I_t).$$

Example I

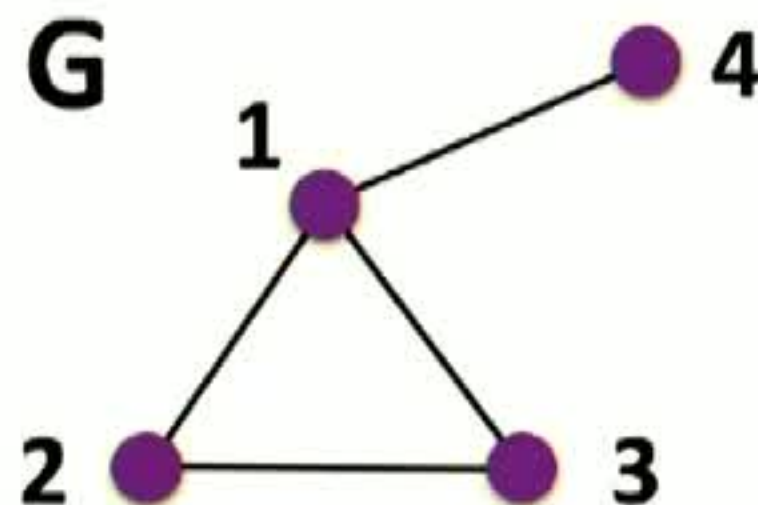
Consider the graph G in the figure below.

- 1 The number of edges of G :
 $t_1 = \sum_{i < j} A_{ij} = 4.$
- 2 The number of 2-paths in G :
 $t_2 = \sum_{i < j} (A^2)_{ij} = 5.$
- 3 The number of triangles in G :
 $t_3 = \frac{1}{6} \text{tr}(A^3) = 1.$

So, $t = T(G) = (4, 5, 1).$

The reference ideal of G is:

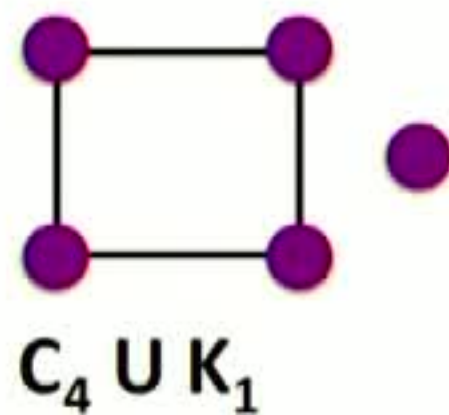
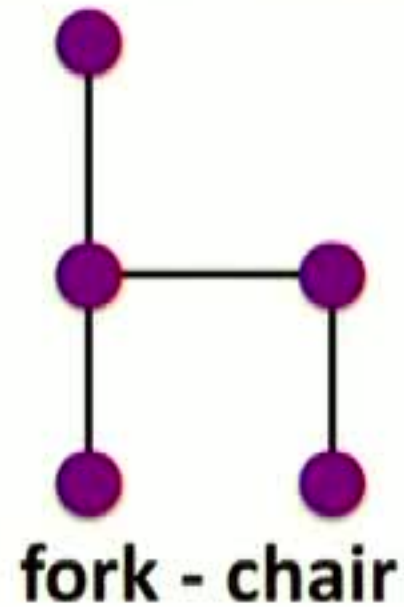
$$\begin{aligned} I_t = \langle & x_{12} + x_{13} + x_{23} + x_{14} + x_{24} + x_{34} - 4, \\ & x_{12}x_{13} + x_{12}x_{23} + x_{13}x_{23} + x_{12}x_{14} + x_{13}x_{14} + x_{14}^2 + x_{12}x_{24} + \\ & + x_{23}x_{24} + x_{24}^2 + x_{13}x_{34} + x_{23}x_{34} + x_{14}x_{34} + x_{24}x_{34} + x_{34}^2 - 5, \\ & x_{12}x_{13}x_{23} + x_{12}x_{14}x_{24} + x_{13}x_{14}x_{34} + x_{23}x_{24}x_{34} - 1 \rangle. \end{aligned}$$



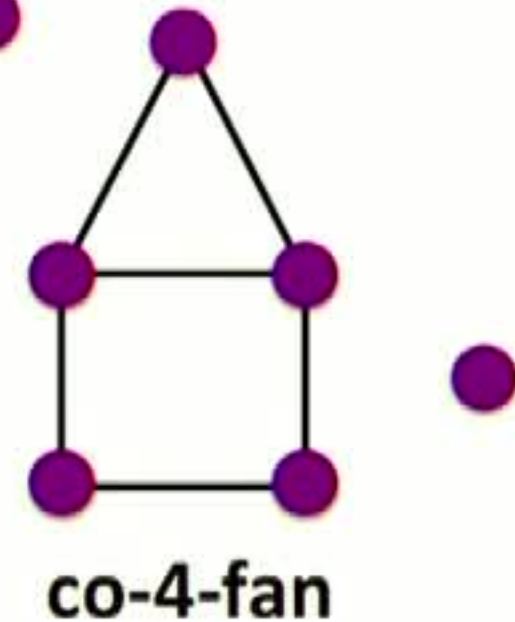
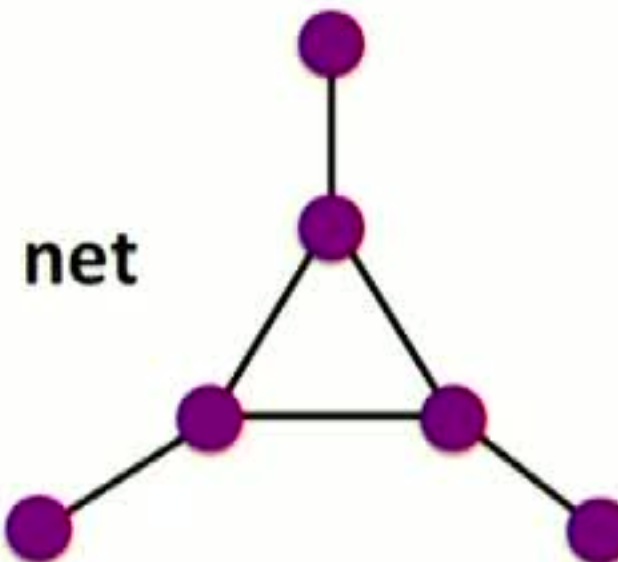
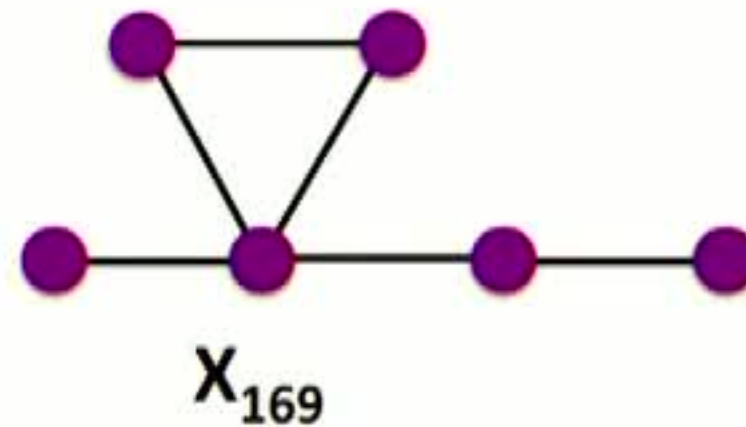
$$\begin{aligned} \text{degree}(I_t) &= 6 \\ \dim(V_t) &= 3 \end{aligned}$$

Example II

• $t_1 = (4, 4, 0)$, V_{t_1} contains:



$t_2 = (6, 9, 1)$, V_{t_2} contains:



Proposition

The polynomials

$$g_1 = f_1,$$

$$g_2 = \left(\sum_{i=2}^n x_{1i} + x_{2i} \right) f_1 - f_2,$$

$$g_3 = f_3 - \left(\sum_{i=3}^n x_{1i} x_{2i} \right) f_1 + x_{23} g_2$$

form a reduced Gröbner basis for I_t . In particular,

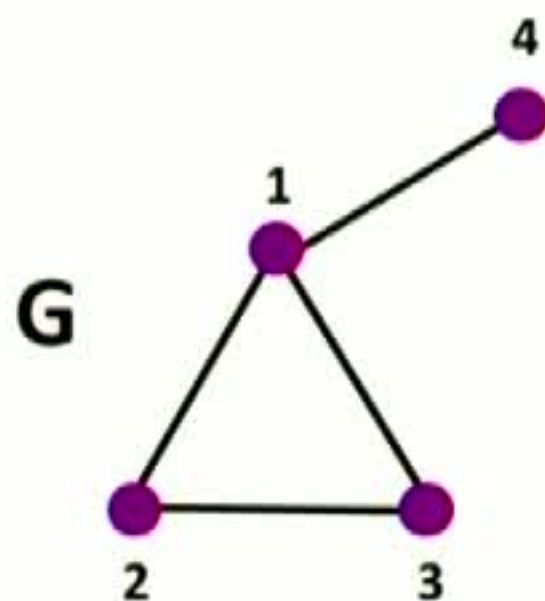
$$In(I_t) = \langle x_{12}, x_{13}^2, x_{23}^3 \rangle.$$

Theorem

• The ideal I_t has codimension 3.

That is, the dimension of the reference variety V_t is:

$$\dim(V_t) = \binom{n}{2} - 3.$$



$$t = T(G) = (4, 5, 1)$$

$$\begin{aligned} \dim(V_t) &= \binom{4}{2} - 3 \\ &= 6 - 3 = 3 \end{aligned}$$

Theorem

If $n \geq 4$, then I_t is prime.

In other words, if $n \geq 4$, then V_t is irreducible.

Proof.

- Fix $h_1, h_2 \notin I_t$ such that $h_1 h_2 \in I_t$. Based on the Gröbner basis, h_1 and h_2 have specific forms.
- $n = 4$: perform the polynomial long division on $h_1 h_2$ by g_2 and g_3 in Macaulay2. The result indicates that either $h_1 = 0$ or $h_2 = 0$.
- $n > 4$: by using division algorithm and induction, we show that the division of $h_1 h_2$ by g_2 and g_3 yields a remainder $r \neq 0$.

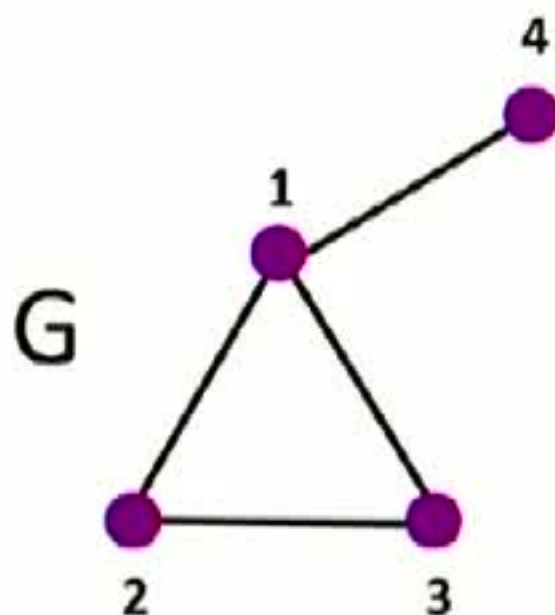


Fiber Singularities

Edge-triangle Model:

Let G be a graph with n vertices. Let $t = T(G) = (t_1, t_3)$. The Jacobian of V_t at G , denoted $Jac(V_t)|_G$, is an $(n^2 - n) \times 2$ matrix whose rows are indexed by pairs (r, s) with $1 \leq r, s \leq n$ and $r \neq s$. The (r, s) th row of $Jac(V_t)|_G$ has the following form

$$[1 \quad \# \text{ of 2-paths between } r \text{ and } s].$$



Rank 2 Jacobian

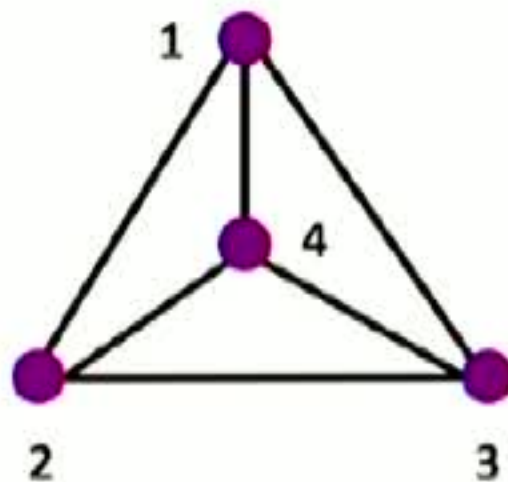
Fiber Singularities

Edge-triangle Model:

Proposition

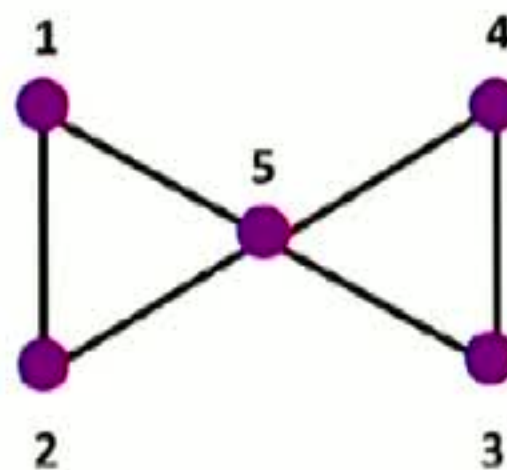
A graph G with n vertices is a singular point of a fiber of the edge-triangle model $\Leftrightarrow G$ is a windmill graph or a strongly regular graph of the form $(n, k, \frac{k^2-k}{n-1}, \frac{k^2-k}{n-1})$.

$n = 4$



K_4

$n = 5$



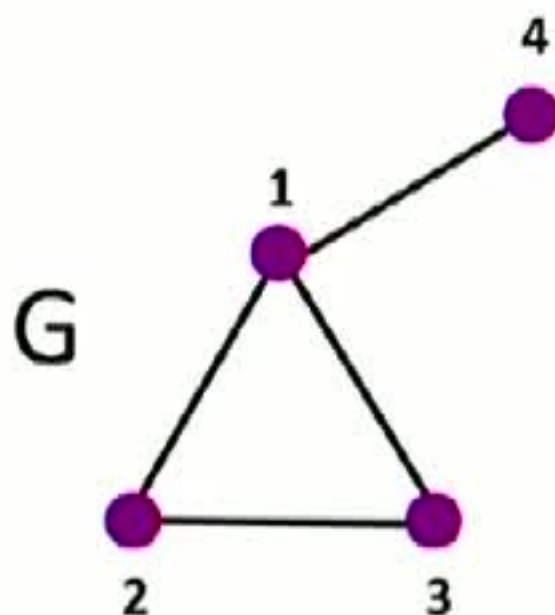
butterfly = hourglass

Fiber Singularities

Edge-triangle Model:

Let G be a graph with n vertices. Let $t = T(G) = (t_1, t_3)$. The Jacobian of V_t at G , denoted $Jac(V_t)|_G$, is an $(n^2 - n) \times 2$ matrix whose rows are indexed by pairs (r, s) with $1 \leq r, s \leq n$ and $r \neq s$. The (r, s) th row of $Jac(V_t)|_G$ has the following form

$$[1 \quad \# \text{ of 2-paths between } r \text{ and } s].$$



Rank 2 Jacobian

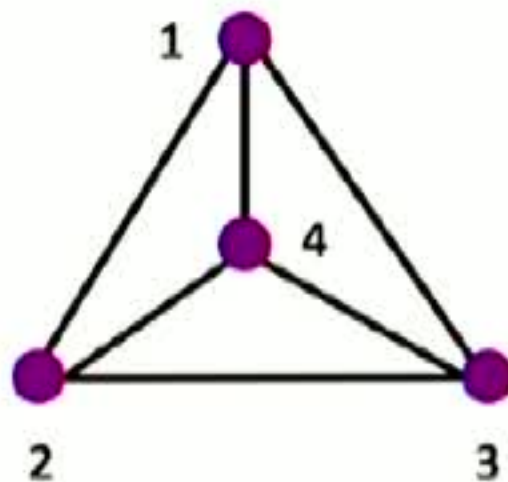
Fiber Singularities

Edge-triangle Model:

Proposition

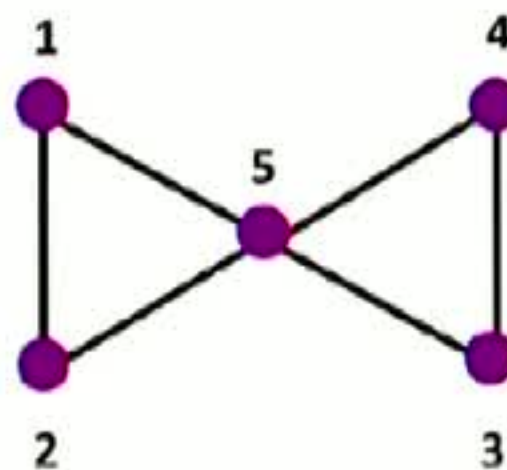
A graph G with n vertices is a singular point of a fiber of the edge-triangle model $\Leftrightarrow G$ is a windmill graph or a strongly regular graph of the form $(n, k, \frac{k^2-k}{n-1}, \frac{k^2-k}{n-1})$.

$n = 4$



K_4

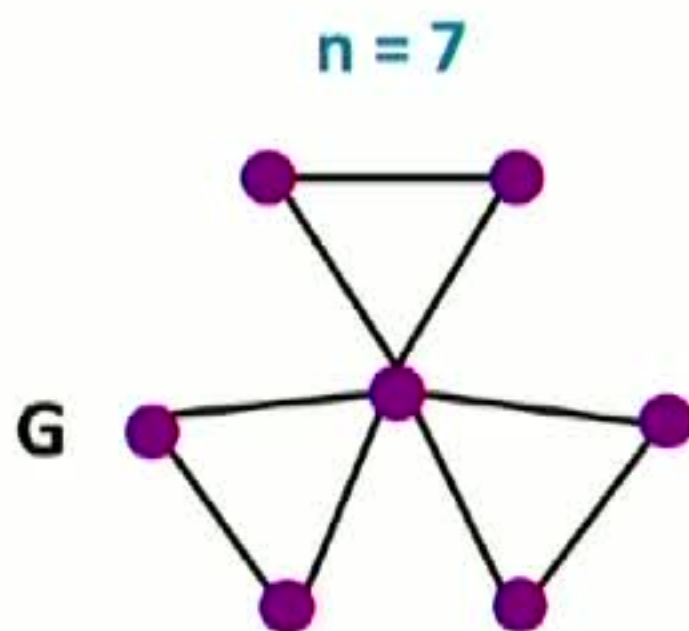
$n = 5$



butterfly = hourglass

Remark

Just because a graph is singular in the edge-triangle model, it does not mean that it is singular in the edge-2path-triangle model.



Edge-triangle Model:

Rank 1

Edge-2path-triangle Model:

Rank 3 (full rank)

Future Directions



- 1 Explore the relationship between singularity and degeneracy.
- 2 Can we use the geometry of the fiber to develop a sampling algorithm for the fiber?

Fiber Singularities

Edge-2path-triangle Model:

Proposition

Let G be a graph. If the automorphism group of G is the full symmetric group, then $\text{rank } \text{Jac}(V_t)|_G = 1$.

Proof.

Proof by contradiction: suppose $\text{rank } \text{Jac}(V_t)|_G \neq 1$.

Case 1: The number of 2-paths between r, s and u, v are different.

Case 2: The joint degrees between r, s and u, v are different.

In both cases, the permutation σ on G is not a graph automorphism (contradiction). Therefore, $\text{Jac}(V_t)|_G = 1$. $\square$

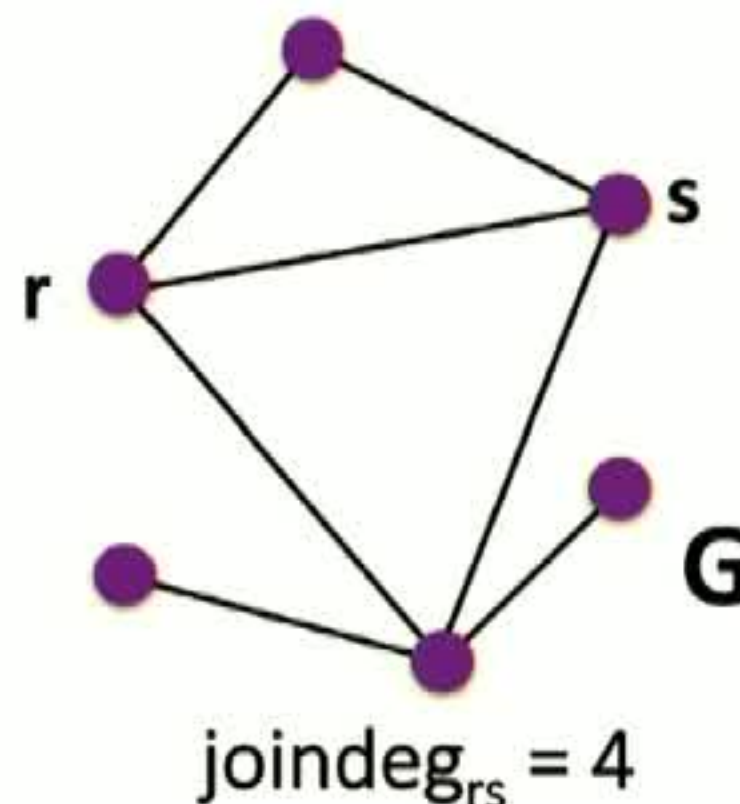


Fiber Singularities

Edge-2path-triangle Model:

Definition

Given a graph $G = (V, E)$ with adjacency matrix A and two vertices $r, s \in V$, let the joint degree of r and s in G be $\text{jointdeg}_{rs} := \deg r + \deg s - 2A_{rs}$.



The Jacobian of V_t at G , denoted $Jac(V_t)|_G$, is an $(n^2 - n) \times 3$ matrix whose rows are indexed by pairs (r, s) with $1 \leq r, s \leq n$ and $r \neq s$. The (r, s) th row of $Jac(V_t)|_G$ has the following form

$$[1 \quad \text{jointdeg}_{rs} \quad \# \text{ of 2-paths between } r \text{ and } s].$$

Fiber Singularities

Edge-2path-triangle Model:

Proposition

Let G be a graph. If the automorphism group of G is the full symmetric group, then $\text{rank } \text{Jac}(V_t)|_G = 1$.

Proof.

Proof by contradiction: suppose $\text{rank } \text{Jac}(V_t)|_G \neq 1$.

Case 1: The number of 2-paths between r, s and u, v are different.

Case 2: The joint degrees between r, s and u, v are different.

In both cases, the permutation σ on G is not a graph automorphism (contradiction). Therefore, $\text{Jac}(V_t)|_G = 1$. $\square$



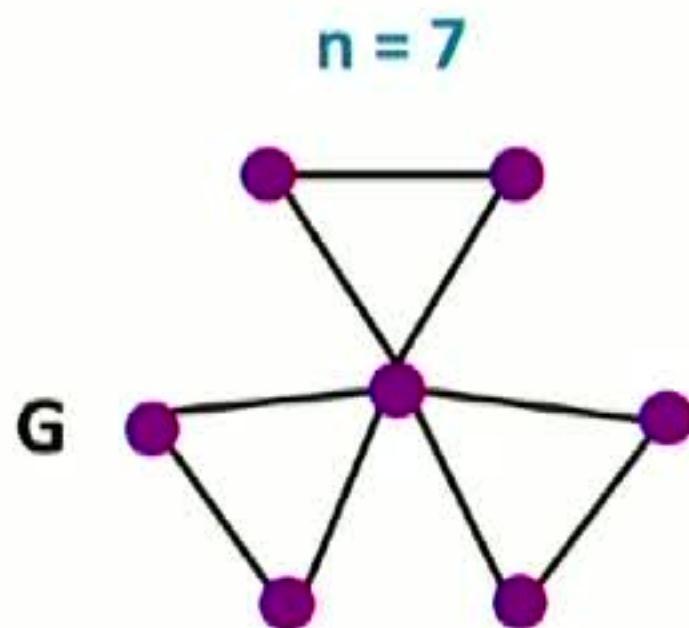
Future Directions



- 1 Explore the relationship between singularity and degeneracy.
- 2 Can we use the geometry of the fiber to develop a sampling algorithm for the fiber?

Remark

Just because a graph is singular in the edge-triangle model, it does not mean that it is singular in the edge-2path-triangle model.



Edge-triangle Model:

Rank 1

Edge-2path-triangle Model:

Rank 3 (full rank)