Identifying large bipartite subgraphs of a graph: combinatorial versus spectral approaches

Debdas Paul

Institute for Systems Theory and Automatic Control
University of Stuttgart, Germany

Collaborator: Prof. Dragan Stevanović
Affiliation: University of Niš, PMF, Serbia and University of Primorska, UP IAM, Slovenia

WGSCO-2018, Aveiro, Portugal, January 27-29

Bipartite graphs have important applications in various fields of science and technology.

Examples

- Decode code words received from the channel (Factor graphs and Tanner graphs)
- Petri nets (Directed bipartite graphs)
- Movies preferences: How much someone would enjoy a movie based on their preferences.
- . . .

In fact it has been shown that *all complex networks can be viewed as bipartite structures sharing some important statistics..*

-Guillaume, J. L., & Latapy, M. (2004), *Info. proc. lett.*, **90**(5)

Reprinted from
ISRAEL JOURNAL OF MATHEMATICS
Volume 3, Number 2, June 1965

ON SOME EXTREMAL PROBLEMS IN GRAPH THEORY

BY
P. ERDÖS

ABSTRACT

The author proves that if C is a sufficiently large constant then every graph of n vertices and $[Cn^{3/2}]$ edges contains a hexagon $X_1, X_2, X_3, X_4, X_5, X_6$ and a seventh vertex Y joined to X_1, X_3 and X_5. The problem is left open whether our graph contains the edges of a cube, (i.e. an eight vertex Z joined to X_2, X_4 and X_6).

Lemma

Every $G(n; m)$ has an even subgraph having at least $m/2$ edges.

Making a graph bipartite

Reprinted from
ISRAEL JOURNAL OF MATHEMATICS
Volume 3, Number 2, June 1965

ON SOME EXTREMAL PROBLEMS IN GRAPH THEORY

BY
P. ERDÖS

ABSTRACT

The author proves that if C is a sufficiently large constant then every graph of n vertices and $[Cn^{3/2}]$ edges contains a hexagon $X_1, X_2, X_3, X_4, X_5, X_6$ and a seventh vertex Y joined to X_1, X_3 and X_5. The problem is left open whether our graph contains the edges of a cube, (i.e. an eight vertex Z joined to X_2, X_4 and X_6).

Lemma

Every $G(n; m)$ has an even subgraph having at least $m/2$ edges.

For a simple, connected graph $G(V, E)$ having a degree matrix D, the three characteristic matrices indicate bipartivity in the following way:

- Adjacency matrix (A)
  - G is bipartite $\implies \lambda^{A}_{\text{min}} = -\lambda^{A}_{\text{max}}$
  - G is non-bipartite, then λ^{A}_{min} is nearer to 0
- Normalized Laplacian matrix ($\mathcal{NL} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, $L = D - A$)
  - G is bipartite $\iff \lambda^{\mathcal{NL}}_{\text{max}} = 2$
- Signless Laplacian matrix ($Q = D + A$)
  - G is bipartite $\iff \lambda^{Q}_{\text{min}} = 0$

Other measure:

$$\beta(G) = \frac{\#\text{even closed walks}}{\text{Total}\,\#\,\text{closed walks}} = \frac{\sum_{i=1}^{N} cosh(\lambda_i)}{\sum_{i=1}^{N} \exp(\lambda_i)}, \lambda_i \in \lambda_A$$

-Estrada, E., & Rodríguez-Velázquez, J. A. (2005), Phys. Rev. E, **72**(4)

For a simple, connected graph $G(V, E)$ having a degree matrix D, the three characteristic matrices indicate bipartivity in the following way:

- Adjacency matrix (A)
  - G is bipartite $\implies \lambda^A_{\text{min}} = -\lambda^A_{\text{max}}$
  - G is non-bipartite, then λ^A_{min} is nearer to 0
- Normalized Laplacian matrix ($\mathcal{NL} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, $L = D - A$)
  - G is bipartite $\iff \lambda^{\mathcal{NL}}_{\text{max}} = 2$
- Signless Laplacian matrix ($Q = D + A$)
  - G is bipartite $\iff \lambda^Q_{\text{min}} = 0$

Other measure:

$$\beta(G) = \frac{\#\text{even closed walks}}{\text{Total}\,\#\,\text{closed walks}} = \frac{\sum_{i=1}^{N} cosh(\lambda_i)}{\sum_{i=1}^{N} \exp(\lambda_i)}, \lambda_i \in \lambda_A$$

-Estrada, E., & Rodríguez-Velázquez, J. A. (2005), Phys. Rev. E, **72**(4)

Spectral graph theory on bipartivity

For a simple, connected graph $G(V, E)$ having a degree matrix D, the three characteristic matrices indicate bipartivity in the following way:

- Adjacency matrix (A)
  - G is bipartite $\implies \lambda^A_{\text{min}} = -\lambda^A_{\text{max}}$
  - G is non-bipartite, then λ^A_{min} is nearer to 0
- Normalized Laplacian matrix ($\mathcal{NL} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}},\ L = D - A$)
  - G is bipartite $\iff \lambda^{\mathcal{NL}}_{\text{max}} = 2$
- Signless Laplacian matrix ($Q = D + A$)
  - G is bipartite $\iff \lambda^Q_{\text{min}} = 0$

Other measure:

$$\beta(G) = \frac{\#\text{even closed walks}}{\text{Total}\,\#\,\text{closed walks}} = \frac{\sum_{i=1}^{N} cosh(\lambda_i)}{\sum_{i=1}^{N} \exp(\lambda_i)}, \lambda_i \in \lambda_A$$

-Estrada, E., & Rodríguez-Velázquez, J. A. (2005), Phys. Rev. E, **72**(4)

For a simple, connected graph $G(V, E)$ having a degree matrix D, the three characteristic matrices indicate bipartivity in the following way:

- Adjacency matrix (A)
  - G is bipartite $\implies \lambda^A_{\text{min}} = -\lambda^A_{\text{max}}$
  - G is non-bipartite, then λ^A_{min} is nearer to 0
- Normalized Laplacian matrix ($\mathcal{NL} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, $L = D - A$)
  - G is bipartite $\iff \lambda^{\mathcal{NL}}_{\text{max}} = 2$
- Signless Laplacian matrix ($Q = D + A$)
  - G is bipartite $\iff \lambda^{Q}_{\text{min}} = 0$

Other measure:

$$\beta(G) = \frac{\#\text{even closed walks}}{\text{Total}\,\#\,\text{closed walks}} = \frac{\sum_{i=1}^{N} cosh(\lambda_i)}{\sum_{i=1}^{N} \exp(\lambda_i)}, \lambda_i \in \lambda_A$$

-Estrada, E., & Rodríguez-Velázquez, J. A. (2005), Phys. Rev. E, **72**(4)

Spectral graph theory on bipartivity

For a simple, connected graph $G(V, E)$ having a degree matrix D, the three characteristic matrices indicate bipartivity in the following way:

- Adjacency matrix (A)
  - G is bipartite $\implies \lambda^A_{\text{min}} = -\lambda^A_{\text{max}}$
  - G is non-bipartite, then λ^A_{min} is nearer to 0
- Normalized Laplacian matrix ($\mathcal{NL} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}}$, $L = D - A$)
  - G is bipartite $\iff \lambda^{\mathcal{NL}}_{\text{max}} = 2$
- Signless Laplacian matrix ($Q = D + A$)
  - G is bipartite $\iff \lambda^Q_{\text{min}} = 0$

Other measure:

$$\beta(G) = \frac{\#\text{even closed walks}}{\text{Total}\,\#\,\text{closed walks}} = \frac{\sum_{i=1}^{N} cosh(\lambda_i)}{\sum_{i=1}^{N} \exp(\lambda_i)}, \lambda_i \in \lambda_A$$

-Estrada, E., & Rodríguez-Velázquez, J. A. (2005), Phys. Rev. E, **72**(4)

Strengths and weaknesses

Matrix	Bipartite	# Components	# Bipartite Components	# Edges
Adjacency	Yes	No	No	Yes
Laplacian	No	Yes	No	Yes
Signless Laplacian	No	No	Yes	Yes
Normalized Laplacian	Yes	Yes	Yes	No

– Butler. S and Chung. F., Handbook of Linear Algebra, 2nd Eds. 2013

A **No** indicates the existence of two non-isomorphic graphs which have the same spectrum but differ in the indicated structure

Examples

- Multiplicity of $\lambda Q = 0$ indicates the $\#$ of bipartite components
- $\#$ of bipartite components is $\alpha + n - \sum_{i=1} \lambda_i^{\mathcal{NL}}$, α is the multiplicity of 2

- How can we find large bipartite subgraphs in G using the argument by Erdös and the spectrum of A, L, $\mathcal{NL}$, and Q?

- How efficient (in terms of size) are those methods in finding large subgraphs?

- How can we find large bipartite subgraphs in G using the argument by Erdös and the spectrum of A, L, $\mathcal{NL}$, and Q?

- How efficient (in terms of size) are those methods in finding large subgraphs?

1. Finding the largest bipartition in $G(V, E)$
   - Combinatorial approach
   - Spectral approaches
   - Initial results
2. New measures
   - Results with new measures
3. Summary and outlook

1. Finding the largest bipartition in $G(V, E)$
   - Combinatorial approach
   - Spectral approaches
   - Initial results

2. New measures

3. Summary and outlook

1. partition V into two disjoint sets $(V_1 \cup V_2)$
2. transfer a vertex $v_i \in V_j$ to $V_{k \neq j}, j, k = \{1, 2\}$ if

$$\frac{2 * \deg(v_i) \text{in } V_j}{deg(v_i)} > 1$$

3. terminate if there is no more new movement.

Input: A simple, connected graph $G = (E, V)$

Result: $\frac{|E_{bipart}(G)|}{|E(G)|}$

Initialize: $V_{final} = \Phi$; $X \subset V$ and $Y = V \backslash \{X\}$

```
repeat
    if S_start = X(or Y) then
        if 2E_u^X > E_u^V ∧ u ∉ V_final (2E_v^Y > E_v^V ∧ v ∉ V_final) then
            Y ← u (X ← v);
            V_final ← u (V_final ← v)
        S_start ← Y (X)
until There is no movement of vertices;
```

$E_i^H := \{|E| \; | i \in H \wedge (\forall j \in H, (i, j) \in E)\}$

Note: To minimize the effect of random partitioning, we consider the average of 10 different partitions.

1. partition V into two disjoint sets $(V_1 \cup V_2)$
2. transfer a vertex $v_i \in V_j$ to $V_{k \neq j}, j, k = \{1, 2\}$ if

$$\frac{2 * \deg(v_i) \text{in } V_j}{deg(v_i)} > 1$$

3. terminate if there is no more new movement.

Input: A simple, connected graph $G = (E, V)$

Result: $\frac{|E_{bipart}(G)|}{|E(G)|}$

Initialize: $V_{final} = \Phi$; $X \subset V$ and $Y = V \backslash \{X\}$

```
repeat
    if S_start = X(or Y) then
        if 2E_u^X > E_u^V ∧ u ∉ V_final
        (2E_v^Y > E_v^V ∧ v ∉ V_final) then
            Y ← u (X ← v);
            V_final ← u (V_final ← v)
        S_start ← Y (X)
until There is no movement of vertices;
```

$E_i^H := \{|E| \ |i \in H \wedge (\forall j \in H, (i, j) \in E)\}$

Note: To minimize the effect of random partitioning, we consider the average of 10 different partitions.

Combinatorial approach

1. partition V into two disjoint sets $(V_1 \cup V_2)$
2. transfer a vertex $v_i \in V_j$ to $V_{k \neq j}, j, k = \{1, 2\}$ if

$$\frac{2 * \deg(v_i) \text{in } V_j}{deg(v_i)} > 1$$

3. terminate if there is no more new movement.

Input: A simple, connected graph $G = (E, V)$
Result: $\frac{|E_{bipart}(G)|}{|E(G)|}$
Initialize: $V_{final} = \Phi$; $X \subset V$ and $Y = V \backslash \{X\}$

```
repeat
    if S_start = X(or Y) then
        if 2E_u^X > E_u^V ∧ u ∉ V_final (2E_v^Y > E_v^V ∧ v ∉ V_final) then
            Y ← u (X ← v);
            V_final ← u (V_final ← v)
        S_start ← Y (X)
until There is no movement of vertices;
```

$E_i^H := \{|E| \; | i \in H \wedge (\forall j \in H, (i, j) \in E)\}$

Note: To minimize the effect of random partitioning, we consider the average of 10 different partitions.

- Eigenvector based approach
  1. extract the non-zero entries of the eigenvectors corresponding to
     - $\lambda_{\min}^{A}$ and $\lambda_{\min}^{Q}$
     - $\lambda_{\max}^{\mathcal{NL}}$ and $\lambda_{\max}^{L}$
  2. make the bipartition based on the signs of the non-zero entries.
- Approach based on Estrada and Rodríguez-Velázquez's $\beta(G)$
  1. $e \leftarrow \arg\min_{e \in E} 1 - [\beta(G-e) - \beta(G)]$
  2. $G \leftarrow G - e$
  3. Repeat steps 1 and 2 until G becomes bipartite.

- Eigenvector based approach
  1. extract the non-zero entries of the eigenvectors corresponding to
     - $\lambda^{A}_{\min}$ and $\lambda^{Q}_{\min}$
     - $\lambda^{\mathcal{NL}}_{\max}$ and $\lambda^{L}_{\max}$
  2. make the bipartition based on the signs of the non-zero entries.
- Approach based on Estrada and Rodríguez-Velázquez's $\beta(G)$
  1. $e \leftarrow \arg\min_{e \in E} 1 - [\beta(G-e) - \beta(G)]$
  2. $G \leftarrow G - e$
  3. Repeat steps 1 and 2 until G becomes bipartite.

Graph Models	Parameter(s)	initial value : step size :final value
Erdös-Rényi (E-R)	p	0.1 : 0.1 : 1
Watts-Strogatz (W-S)	(β, k)	(0.3, 1 : 1 : 9)
Barabási-Albert (B-A)	n	1 : 1 : 10

- p: Probabilty of attachment
- (β, k): probability of rewiring and the mean degree
- n: number of edges to attach in every step

We generate 1000 different graphs with 20 vertices corresponding to each values in the respective model parameters.

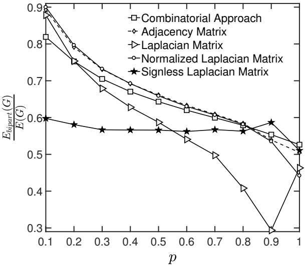
Choice of graph models

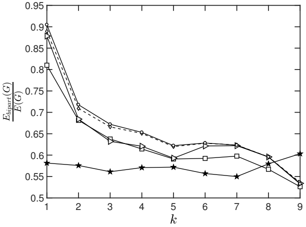
Graph Models	Parameter(s)	initial value : step size :final value
Erdös-Rényi (E-R)	p	0.1 : 0.1 : 1
Watts-Strogatz (W-S)	(β, k)	(0.3, 1 : 1 : 9)
Barabási-Albert (B-A)	n	1 : 1 : 10

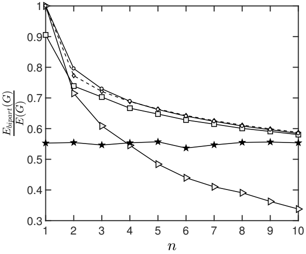
- p: Probabilty of attachment
- (β, k): probability of rewiring and the mean degree
- n: number of edges to attach in every step

We generate 1000 different graphs with 20 vertices corresponding to each values in the respective model parameters.

Eigenvector vs. combinatorial approach



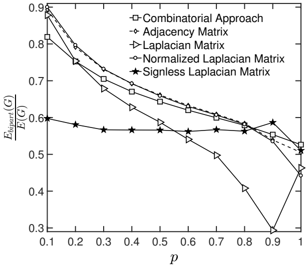


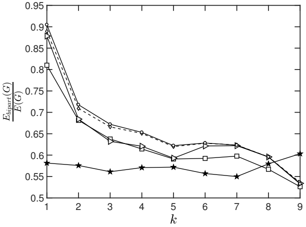


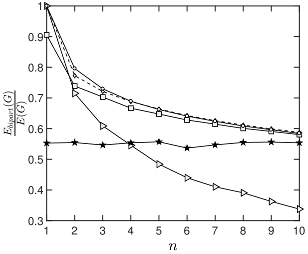
Remark:

- Performances of $\mathcal{NL}$ and A matrices are comparable even slightly better than that of the combinatorial approach.

Eigenvector vs. combinatorial approach







Remark:

- Performances of $\mathcal{NL}$ and A matrices are comparable even slightly better than that of the combinatorial approach.

1. Finding the largest bipartition in $G(V, E)$

2. New measures
   - Results with new measures

3. Summary and outlook

Let, Spectrum of $A(G) = (\boldsymbol{\nu}, \boldsymbol{\lambda})$
$\boldsymbol{\nu}$:= set of eigenvectors
$\boldsymbol{\lambda}$:= set of eigenvalues$=\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|V(G)|}\}$

- A matrix based approach
  1. $e \leftarrow \arg\max_{e \in E(G)} \frac{\nu^i_{\lambda_{|V(G)|}} \times \nu^j_{\lambda_{|V(G)|}}}{\nu^i_{\lambda_1} \times \nu^j_{\lambda_1}}$
  2. $G \leftarrow (G - e)$
- $\mathcal{NL}$ matrix based approach
  1. $e \leftarrow \arg\max_{e \in E(G)} \frac{\nu^i_{\gamma_1} \times \nu^j_{\gamma_1}}{\nu^i_{\gamma_{|V(G)|}} \times \nu^j_{\gamma_{|V(G)|}}}$
  2. $G \leftarrow (G - e)$

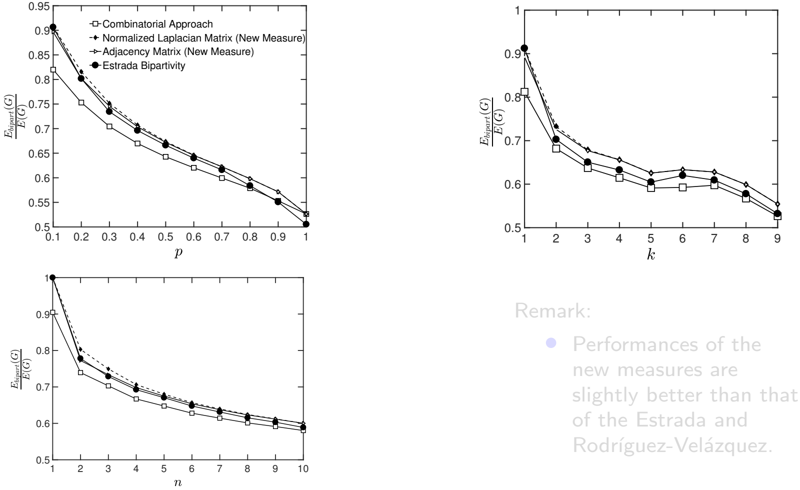
Let, Spectrum of $A(G) = (\boldsymbol{\nu}, \boldsymbol{\lambda})$
$\boldsymbol{\nu}$:= set of eigenvectors
$\boldsymbol{\lambda}$:= set of eigenvalues$=\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|V(G)|}\}$

- A matrix based approach
  1. $e \leftarrow \arg\max_{e \in E(G)} \frac{\nu^i_{\lambda_{|V(G)|}} \times \nu^j_{\lambda_{|V(G)|}}}{\nu^i_{\lambda_1} \times \nu^j_{\lambda_1}}$
  2. $G \leftarrow (G - e)$
- $\mathcal{NL}$ matrix based approach
  1. $e \leftarrow \arg\max_{e \in E(G)} \frac{\nu^i_{\gamma_1} \times \nu^j_{\gamma_1}}{\nu^i_{\gamma_{|V(G)|}} \times \nu^j_{\gamma_{|V(G)|}}}$
  2. $G \leftarrow (G - e)$

Let, Spectrum of $A(G) = (\boldsymbol{\nu}, \boldsymbol{\lambda})$
$\boldsymbol{\nu}$:= set of eigenvectors
$\boldsymbol{\lambda}$:= set of eigenvalues$=\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|V(G)|}\}$

- A matrix based approach
  1. $e \leftarrow \arg\max_{e \in E(G)} \frac{\nu^i_{\lambda_{|V(G)|}} \times \nu^j_{\lambda_{|V(G)|}}}{\nu^i_{\lambda_1} \times \nu^j_{\lambda_1}}$
  2. $G \leftarrow (G - e)$
- $\mathcal{NL}$ matrix based approach
  1. $e \leftarrow \arg\max_{e \in E(G)} \frac{\nu^i_{\gamma_1} \times \nu^j_{\gamma_1}}{\nu^i_{\gamma_{|V(G)|}} \times \nu^j_{\gamma_{|V(G)|}}}$
  2. $G \leftarrow (G - e)$

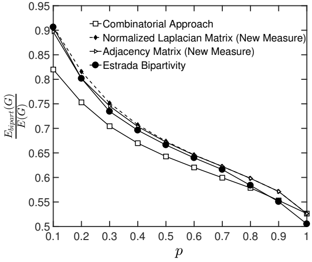
Performances of the new measures

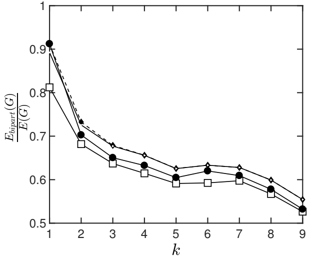


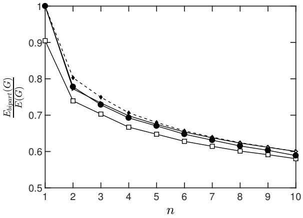
Remark:

- Performances of the new measures are slightly better than that of the Estrada and Rodríguez-Velázquez.

Performances of the new measures







Remark:

- Performances of the new measures are slightly better than that of the Estrada and Rodríguez-Velázquez.

1. Finding the largest bipartition in $G(V, E)$
2. New measures
3. Summary and outlook

- Summary
  1. **Aim**: Identification of large bipartite subgraphs of a graph
  2. **Approaches**:
     - Combinatorial approach due to Erdös
     - Spectral approaches involving A, L, $\mathcal{NL}$ and Q matrices
     - Bipartivity $\beta(G)$proposed by Estrada and Rodríguez-Velázquez in 2005
     - New measure of edge bipartivity using A, $\mathcal{NL}$
  3. **Graph models**: E-R, W-S, and B-A
  4. **Preliminary observation**: A, $\mathcal{NL}$ based approaches performs better than combinatorial approach and the approach based on $\beta(G)$ by Estrada and Rodríguez-Velázquez.
- Outlook
  1. Mathematical proofs to justify numerical observations.
  2. Consider other classes of graphs, e.g.- 3-connected cubic planar triangle-free graphs for which the lower bound is $\frac{39n}{32} - \frac{9}{12}$ - Cui. Q, Wang. J, (2009), Discr. Math. 309 (**5**)

- Summary
  1. **Aim**: Identification of large bipartite subgraphs of a graph
  2. **Approaches**:
     - Combinatorial approach due to Erdös
     - Spectral approaches involving A, L, $\mathcal{NL}$ and Q matrices
     - Bipartivity $\beta(G)$proposed by Estrada and Rodríguez-Velázquez in 2005
     - New measure of edge bipartivity using A, $\mathcal{NL}$
  3. **Graph models**: E-R, W-S, and B-A
  4. **Preliminary observation**: A, $\mathcal{NL}$ based approaches performs better than combinatorial approach and the approach based on $\beta(G)$ by Estrada and Rodríguez-Velázquez.
- Outlook
  1. Mathematical proofs to justify numerical observations.
  2. Consider other classes of graphs, e.g.- 3-connected cubic planar triangle-free graphs for which the lower bound is $\frac{39n}{32} - \frac{9}{12}$ - Cui. Q, Wang. J, (2009), Discr. Math. 309 (5)

- Summary
  1. **Aim**: Identification of large bipartite subgraphs of a graph
  2. **Approaches**:
     - Combinatorial approach due to Erdös
     - Spectral approaches involving A, L, $\mathcal{NL}$ and Q matrices
     - Bipartivity $\beta(G)$proposed by Estrada and Rodríguez-Velázquez in 2005
     - New measure of edge bipartivity using A, $\mathcal{NL}$
  3. **Graph models**: E-R, W-S, and B-A
  4. **Preliminary observation**: A, $\mathcal{NL}$ based approaches performs better than combinatorial approach and the approach based on $\beta(G)$ by Estrada and Rodríguez-Velázquez.
- Outlook
  1. Mathematical proofs to justify numerical observations.
  2. Consider other classes of graphs, e.g.- 3-connected cubic planar triangle-free graphs for which the lower bound is $\frac{39n}{32} - \frac{9}{12}$ - Cui. Q, Wang. J, (2009), Discr. Math. 309 (**5**)

- Summary
  1. **Aim**: Identification of large bipartite subgraphs of a graph
  2. **Approaches**:
     - Combinatorial approach due to Erdös
     - Spectral approaches involving A, L, $\mathcal{NL}$ and Q matrices
     - Bipartivity $\beta(G)$proposed by Estrada and Rodríguez-Velázquez in 2005
     - New measure of edge bipartivity using A, $\mathcal{NL}$
  3. **Graph models**: E-R, W-S, and B-A
  4. **Preliminary observation**: A, $\mathcal{NL}$ based approaches performs better than combinatorial approach and the approach based on $\beta(G)$ by Estrada and Rodríguez-Velázquez.
- Outlook
  1. Mathematical proofs to justify numerical observations.
  2. Consider other classes of graphs, e.g.- 3-connected cubic planar triangle-free graphs for which the lower bound is $\frac{39n}{32} - \frac{9}{12}$ - Cui. Q, Wang. J, (2009), Discr. Math. 309 (**5**)

Summary and outlook

- Summary
  1. **Aim**: Identification of large bipartite subgraphs of a graph
  2. **Approaches**:
     - Combinatorial approach due to Erdös
     - Spectral approaches involving A, L, $\mathcal{NL}$ and Q matrices
     - Bipartivity $\beta(G)$proposed by Estrada and Rodríguez-Velázquez in 2005
     - New measure of edge bipartivity using A, $\mathcal{NL}$
  3. **Graph models**: E-R, W-S, and B-A
  4. **Preliminary observation**: A, $\mathcal{NL}$ based approaches performs better than combinatorial approach and the approach based on $\beta(G)$ by Estrada and Rodríguez-Velázquez.
- Outlook
  1. Mathematical proofs to justify numerical observations.
  2. Consider other classes of graphs, e.g.- 3-connected cubic planar triangle-free graphs for which the lower bound is $\frac{39n}{32} - \frac{9}{12}$ - Cui. Q, Wang. J, (2009), Discr. Math. 309 (**5**)

Acknowledgement

The Ministry of Education, Science, Culture and Sport of the Republic of Slovenia

Thank you for your kind attention!