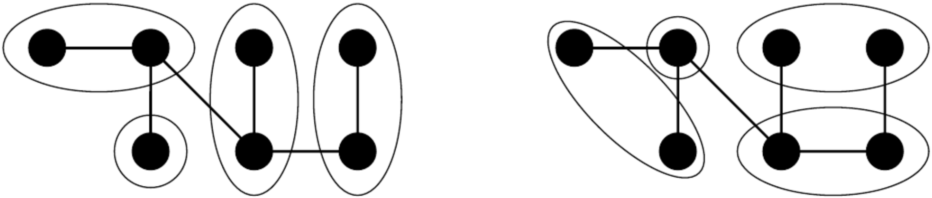
Bounds for the Number of Partitions of a Graph

ECDA 2019 – Bayreuth

Fabian Ball and Andreas Geyer-Schulz | March 18, 2019

INFORMATION SERVICES AND ELECTRONIC MARKETS



Outline

1. Motivation

2. Preliminaries

3. Lower Bound

4. Small Graphs

5. Estimated Number of *k*-Partitions

6. Conclusion

1 Motivation

- We all know the following combinatorial results
  1. The number of k-partitions of an n-set is the Stirling number of the 2nd kind $S(n, k)$ (or $\left\{ {n \atop k} \right\}$)
  2. The number of partitions of an n-set is the Bell number, which is the sum over all $1 \leq k \leq n$ Stirling numbers (2nd kind)
- Both numbers grow rapidly with increasing n (and "intermediate" k)

- But what about graph clustering?
- Are all partitions actually part of the solution space?

$\Rightarrow$ Independent of the optimization goal, we normally only want connected nodes within clusters

- Donath and Hoffman (1973) present lower bounds for a special application (cluster sizes are restricted to be smaller than a threshold)

1 Motivation

- We all know the following combinatorial results
  1. The number of k-partitions of an n-set is the Stirling number of the 2nd kind $S(n,k)$ (or $\left\{ {n \atop k} \right\}$)
  2. The number of partitions of an n-set is the Bell number, which is the sum over all $1 \leq k \leq n$ Stirling numbers (2nd kind)
- Both numbers grow rapidly with increasing n (and "intermediate" k)

- But what about graph clustering?
- Are all partitions actually part of the solution space?

$\Rightarrow$ Independent of the optimization goal, we normally only want connected nodes within clusters

- Donath and Hoffman (1973) present lower bounds for a special application (cluster sizes are restricted to be smaller than a threshold)

1 Motivation

- We all know the following combinatorial results
  1. The number of k-partitions of an n-set is the Stirling number of the 2nd kind $S(n,k)$ (or $\left\{ {n \atop k} \right\}$)
  2. The number of partitions of an n-set is the Bell number, which is the sum over all $1 \leq k \leq n$ Stirling numbers (2nd kind)
- Both numbers grow rapidly with increasing n (and "intermediate" k)

- But what about graph clustering?
- Are all partitions actually part of the solution space?

$\Rightarrow$ Independent of the optimization goal, we normally only want connected nodes within clusters

- Donath and Hoffman (1973) present lower bounds for a special application (cluster sizes are restricted to be smaller than a threshold)

Karlsruhe Institute of Technology

Partition of a Set

- Set M of some entities (e. g. data points)
- Partition $P(M) = \{M_1, \ldots, M_k\}$ into subsets/cells M_i
  - Complete: $\bigcup_i M_i = V$
  - Disjoint: $\forall i \neq j : M_i \cap M_j = \emptyset$
  - No empty subsets: $\forall i : M_i \neq \emptyset$

Number of Partitions

- "How many possibilities exist to divide n elements into a partition of $k \leq n$ subsets?"
- Equivalent to the question: "How large is the search space of kmeans?" (for given k)
- Answer: $S(n, k)$ (the Stirling numbers of the 2nd kind) (e. g. Rennie and Dobson, 1969)

Partition of a Set

- Set M of some entities (e. g. data points)
- Partition $P(M) = \{M_1, \ldots, M_k\}$ into subsets/cells M_i
  - Complete: $\bigcup_i M_i = V$
  - Disjoint: $\forall i \neq j : M_i \cap M_j = \emptyset$
  - No empty subsets: $\forall i : M_i \neq \emptyset$

Number of Partitions

- "How many possibilities exist to divide n elements into a partition of $k \leq n$ subsets?"
- Equivalent to the question: "How large is the search space of kmeans?" (for given k)
- Answer: $S(n, k)$ (the Stirling numbers of the 2nd kind) (e. g. Rennie and Dobson, 1969)

Stirling Numbers of the 2nd Kind

- Defined recursively (non-recursive form exists, but not of importance here)
- $S(n,k) = k \cdot S(n-1,k) + S(n-1,k-1)$
- $S(n,1) = S(n,n) = 1$
- Explanation:
  - Shift perspective, add a new element (the $n+1$th)
  - $S(n+1,k) = k \cdot S(n,k) + S(n,k-1)$
  - Either put the new element into one of the k existing cells; for each choice there exist $S(n,k)$ possibilities
  - Or put the element into a new cell itself (the kth one), all other n elements must then be partitioned into $k-1$ cells
- The sum over all $1 \leq k \leq n$ is the Bell number $B(n)$ (Bell, 1934)

Stirling Numbers of the 2nd Kind

- Defined recursively (non-recursive form exists, but not of importance here)
- $S(n,k) = k \cdot S(n-1,k) + S(n-1,k-1)$
- $S(n,1) = S(n,n) = 1$
- Explanation:
  - Shift perspective, add a new element (the $n+1$th)
  - $S(n+1,k) = k \cdot S(n,k) + S(n,k-1)$
  - Either put the new element into one of the k existing cells; for each choice there exist $S(n,k)$ possibilities
  - Or put the element into a new cell itself (the kth one), all other n elements must then be partitioned into $k-1$ cells
- The sum over all $1 \le k \le n$ is the Bell number $B(n)$ (Bell, 1934)

Stirling Numbers of the 2nd Kind

- Defined recursively (non-recursive form exists, but not of importance here)
- $S(n, k) = k \cdot S(n-1, k) + S(n-1, k-1)$
- $S(n, 1) = S(n, n) = 1$
- Explanation:
  - Shift perspective, add a new element (the $n+1$th)
  - $S(n+1, k) = k \cdot S(n, k) + S(n, k-1)$
  - Either put the new element into one of the k existing cells; for each choice there exist $S(n, k)$ possibilities
  - Or put the element into a new cell itself (the kth one), all other n elements must then be partitioned into $k-1$ cells
- The sum over all $1 \leq k \leq n$ is the Bell number $B(n)$ (Bell, 1934)

2 Preliminaries II

Stirling Numbers of the 2nd Kind

- Defined recursively (non-recursive form exists, but not of importance here)
- $S(n,k) = k \cdot S(n-1,k) + S(n-1,k-1)$
- $S(n,1) = S(n,n) = 1$
- Explanation:
  - Shift perspective, add a new element (the $n+1$th)
  - $S(n+1,k) = k \cdot S(n,k) + S(n,k-1)$
  - Either put the new element into one of the k existing cells; for each choice there exist $S(n,k)$ possibilities
  - Or put the element into a new cell itself (the kth one), all other n elements must then be partitioned into $k-1$ cells
- The sum over all $1 \leq k \leq n$ is the Bell number $B(n)$ (Bell, 1934)

Karlsruhe Institute of Technology

Stirling Numbers of the 2nd Kind

- Defined recursively (non-recursive form exists, but not of importance here)
- $S(n, k) = k \cdot S(n-1, k) + S(n-1, k-1)$
- $S(n, 1) = S(n, n) = 1$
- Explanation:
  - Shift perspective, add a new element (the $n + 1$th)
  - $S(n+1, k) = k \cdot S(n, k) + S(n, k-1)$
  - Either put the new element into one of the k existing cells; for each choice there exist $S(n, k)$ possibilities
  - Or put the element into a new cell itself (the kth one), all other n elements must then be partitioned into $k - 1$ cells
- The sum over all $1 \leq k \leq n$ is the Bell number $B(n)$ (Bell, 1934)

2 Preliminaries IV

Simple Graph

- $G = (V, E)$
- Node set $V = \{1, 2, \ldots, n\}$ $(n < \infty)$
- Edge set $E \subseteq \{\{u, v\} \mid u, v \in V,\ u \neq v\}$

$\Rightarrow$ Undirected, finite, no loops, no weights, no multiple edges + connected

Graph Partition

- Defined as partitions above
- The set to be partitioned is the node set V
- **Assumption:** The subgraph induced by a cell is connected (i. e. we never have nodes in the same cell if they are not connected by a path within the cell)

Simple Graph

- $G = (V, E)$
- Node set $V = \{1, 2, \ldots, n\}$ $(n < \infty)$
- Edge set $E \subseteq \{\{u, v\} \mid u, v \in V,\ u \neq v\}$

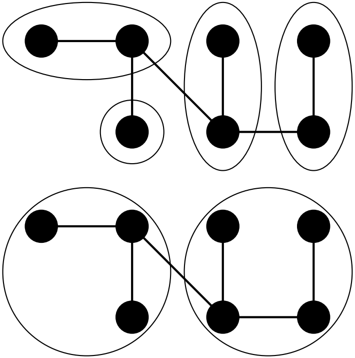
$\Rightarrow$ Undirected, finite, no loops, no weights, no multiple edges + connected

Graph Partition

- Defined as partitions above
- The set to be partitioned is the node set V
- **Assumption:** The subgraph induced by a cell is connected (i. e. we never have nodes in the same cell if they are not connected by a path within the cell)

2 Preliminaries – Graph Partition Assumption Examples

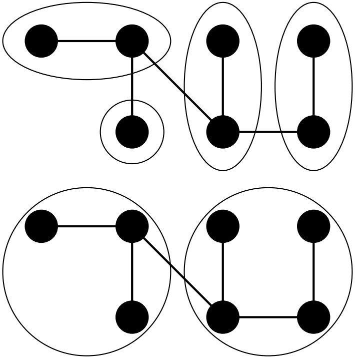
"Valid" Partitions



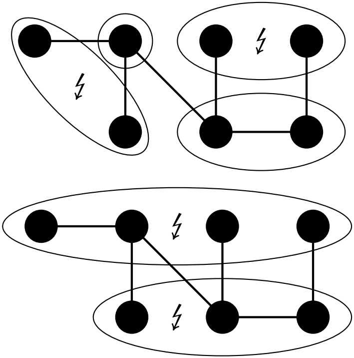
"Invalid" Partitions

2 Preliminaries – Graph Partition Assumption Examples

"Valid" Partitions

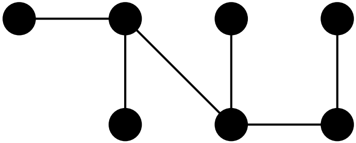


"Invalid" Partitions



3 Lower Bound I

Fact: A tree of n nodes and $m = n - 1$ edges is the sparsest possible connected graph



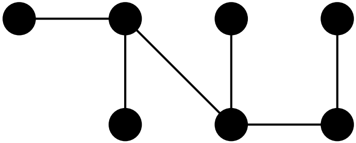
A tree of $n = 7$ nodes

Theorem (Number of Partitions of a Tree)

The number of possible partitions of a tree is $2^m = 2^{n-1}$.

3 Lower Bound I

Fact: A tree of n nodes and $m = n - 1$ edges is the sparsest possible connected graph



A tree of $n = 7$ nodes

Theorem (Number of Partitions of a Tree)

The number of possible partitions of a tree is $2^m = 2^{n-1}$.

Proof.

By induction.

n = 1 $2^{1-1} = 1$, the trivial and the singleton partition are the same

n = 2 $2^{2-1} = 2$, there is the trivial and the singleton partition

n = 3 $2^{3-1} = 4$, trivial + singleton + $2\times$ connected pairs

n + 1 $2^{(n+1)-1} = 2^n = 2 \cdot 2^{n-1} = \underbrace{2^{n-1}}_{(a)} + \underbrace{2^{n-1}}_{(b)}$

The newly added node is connected to exactly one neighbor, and

(a) is added as a singleton cell or

(b) is put into the same cell as its neighbor.

□

Proof.

By induction.

n $= 1$ $2^{1-1} = 1$, the trivial and the singleton partition are the same

n $= 2$ $2^{2-1} = 2$, there is the trivial and the singleton partition

n $= 3$ $2^{3-1} = 4$, trivial + singleton + $2\times$ connected pairs

n $+ 1$ $2^{(n+1)-1} = 2^n = 2 \cdot 2^{n-1} = \underbrace{2^{n-1}}_{(a)} + \underbrace{2^{n-1}}_{(b)}$

The newly added node is connected to exactly one neighbor, and

(a) is added as a singleton cell or

(b) is put into the same cell as its neighbor.

$\square$

Proof.

By induction.

n = **1** $2^{1-1} = 1$, the trivial and the singleton partition are the same

n = **2** $2^{2-1} = 2$, there is the trivial and the singleton partition

n = **3** $2^{3-1} = 4$, trivial + singleton + $2\times$ connected pairs

n + **1** $2^{(n+1)-1} = 2^n = 2 \cdot 2^{n-1} = \underbrace{2^{n-1}}_{\text{(a)}} + \underbrace{2^{n-1}}_{\text{(b)}}$

The newly added node is connected to exactly one neighbor, and

(a) is added as a singleton cell or

(b) is put into the same cell as its neighbor.

□

Proof.

By induction.

n = **1** $2^{1-1} = 1$, the trivial and the singleton partition are the same

n = **2** $2^{2-1} = 2$, there is the trivial and the singleton partition

n = **3** $2^{3-1} = 4$, trivial + singleton + 2× connected pairs

n + **1** $2^{(n+1)-1} = 2^n = 2 \cdot 2^{n-1} = \underbrace{2^{n-1}}_{\text{(a)}} + \underbrace{2^{n-1}}_{\text{(b)}}$

The newly added node is connected to exactly one neighbor, and

(a) is added as a singleton cell or

(b) is put into the same cell as its neighbor.

□

3 Lower Bound II

Karlsruhe Institute of Technology

Proof.

By induction.

$\mathbf{n} = \mathbf{1}$ $2^{1-1} = 1$, the trivial and the singleton partition are the same

$\mathbf{n} = \mathbf{2}$ $2^{2-1} = 2$, there is the trivial and the singleton partition

$\mathbf{n} = \mathbf{3}$ $2^{3-1} = 4$, trivial + singleton + 2× connected pairs

$$\mathbf{n}+\mathbf{1} \quad 2^{(n+1)-1} = 2^n = 2 \cdot 2^{n-1} = \underbrace{2^{n-1}}_{\text{(a)}} + \underbrace{2^{n-1}}_{\text{(b)}}$$

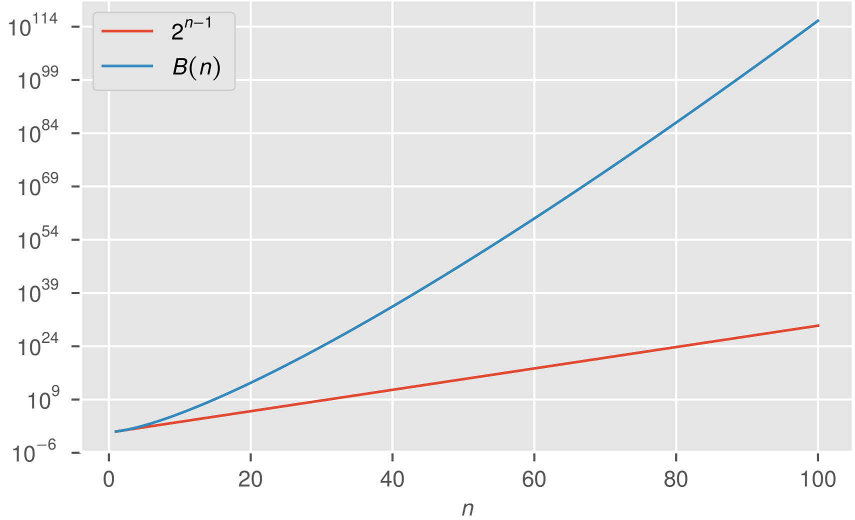
The newly added node is connected to exactly one neighbor, and

(a) is added as a singleton cell or

(b) is put into the same cell as its neighbor.

$\square$

3 Lower Bound vs. Upper Bound (Partitions)



Theorem (Number of k-Partitions of a Tree)

The number of possible partitions of a tree into k cells is $\binom{n-1}{k-1}$.

Proof.

- Two cells are formed by picking (removing) an edge that connects the two subtrees
- $1 \leq k \leq n$ cells are formed by picking $k-1$ unique edges
- The number of choosing $k-1$ elements of from a total of $n-1 = m$ elements is just $\binom{n-1}{k-1}$

□

Karlsruhe Institute of Technology

Theorem (Number of k-Partitions of a Tree)

The number of possible partitions of a tree into k cells is $\binom{n-1}{k-1}$.

Proof.

- Two cells are formed by picking (removing) an edge that connects the two subtrees
- $1 \le k \le n$ cells are formed by picking $k - 1$ unique edges
- The number of choosing $k - 1$ elements of from a total of $n - 1 = m$ elements is just $\binom{n-1}{k-1}$

$\square$

Theorem (Number of k-Partitions of a Tree)

The number of possible partitions of a tree into k cells is $\binom{n-1}{k-1}$.

Proof.

- Two cells are formed by picking (removing) an edge that connects the two subtrees
- $1 \leq k \leq n$ cells are formed by picking $k - 1$ unique edges
- The number of choosing $k - 1$ elements of from a total of $n - 1 = m$ elements is just $\binom{n-1}{k-1}$

$\square$

Alternate Proof for the Number of Partitions

- Using the binomial theorem $(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$
- For $x = y = 1$:

$$2^n = \sum_{k=0}^{n} \binom{n}{k} = \sum_{k=1}^{n+1} \binom{n}{k-1} \tag{1}$$

$$2^{n-1} = \sum_{k=1}^{n} \binom{n-1}{k-1} \tag{2}$$

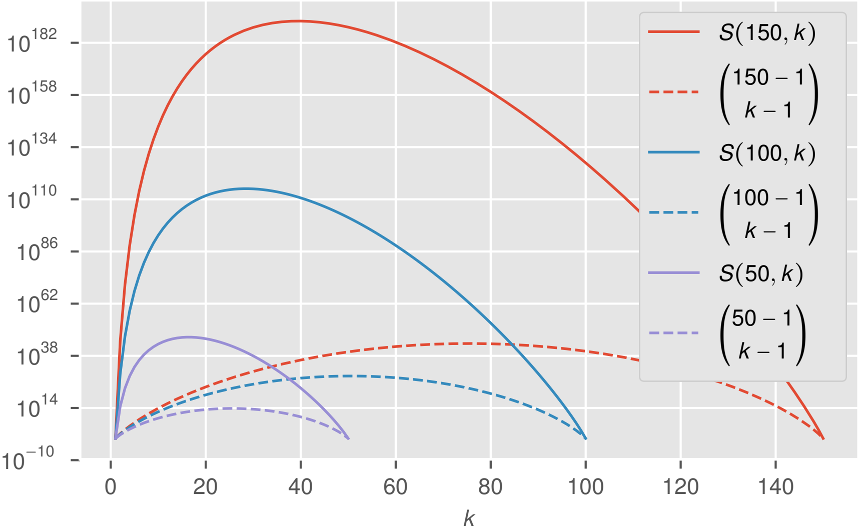
Alternate Proof for the Number of Partitions

- Using the binomial theorem $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$
- For $x = y = 1$:

$$2^n = \sum_{k=0}^{n} \binom{n}{k} = \sum_{k=1}^{n+1} \binom{n}{k-1} \tag{1}$$

$$2^{n-1} = \sum_{k=1}^{n} \binom{n-1}{k-1} \tag{2}$$

3 Lower Bound vs. Upper Bound (k-Partitions)



Conclusion

- The number of clustering partitions of a tree are the exact lower bound for the number of partitions
- Still exponential, but had to be expected (argument: graph clustering is $\mathcal{NP}$-hard)

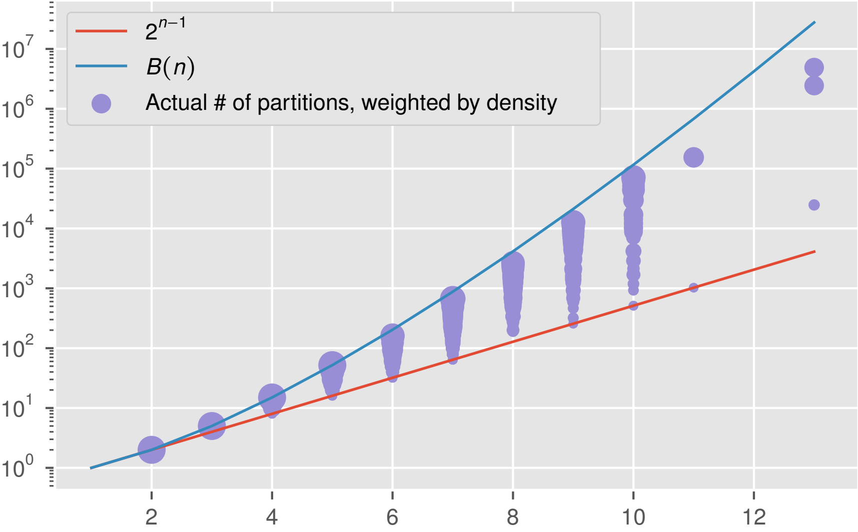
4 Small Graphs – Overview

- Small connected graphs with $n = 2, \dots, 13$ nodes
- Downloaded from `http://www.graphclasses.org/smallgraphs.html`
- Includes *all* graphs with $n = 2, \dots, 5$ nodes and some graphs with $n = 6, \dots, 13$
- Total number of investigated graphs is 501
- We enumerated the *full* search space (given our assumption of connected nodes within clusters) for all these graphs
- Already (very) time and memory consuming for "larger" n (in the range of several minutes and more than 40 GB)
- Graph with the largest search space: $\overline{X_{196}}$ ($n = 13$; $m = 39$; # Partitions is 4,880,943, but $B(13) = 27{,}644{,}437$)

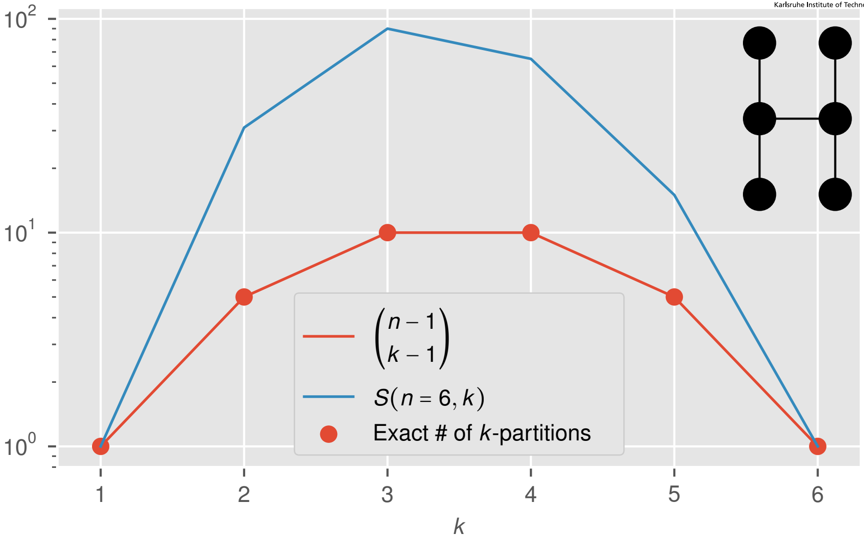
4 Small Graphs – Overview

- Small connected graphs with $n = 2, \ldots, 13$ nodes
- Downloaded from `http://www.graphclasses.org/smallgraphs.html`
- Includes *all* graphs with $n = 2, \ldots, 5$ nodes and some graphs with $n = 6, \ldots, 13$
- Total number of investigated graphs is 501
- We enumerated the *full* search space (given our assumption of connected nodes within clusters) for all these graphs
- Already (very) time and memory consuming for "larger" n (in the range of several minutes and more than 40 GB)
- Graph with the largest search space: $\overline{X_{196}}$ ($n = 13$; $m = 39$; # Partitions is 4,880,943, but $B(13) = 27{,}644{,}437$)

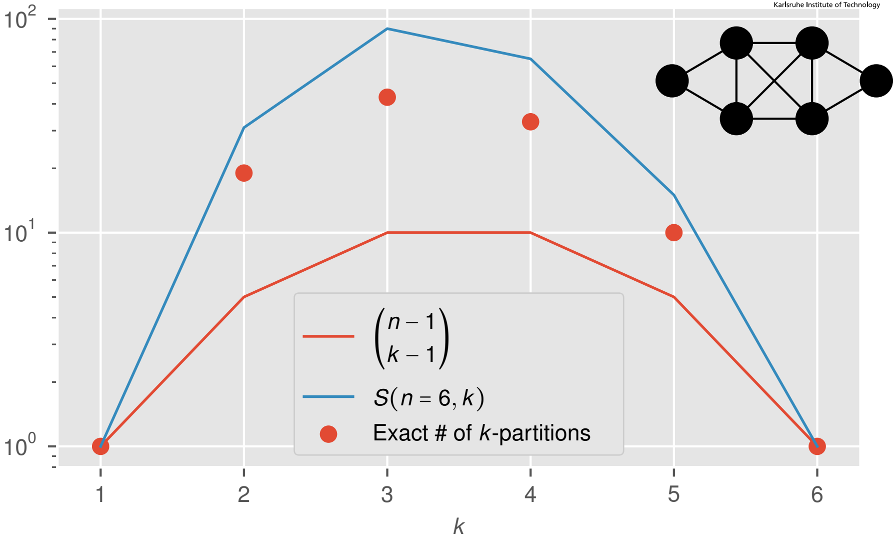
4 Small Graphs – All Graphs



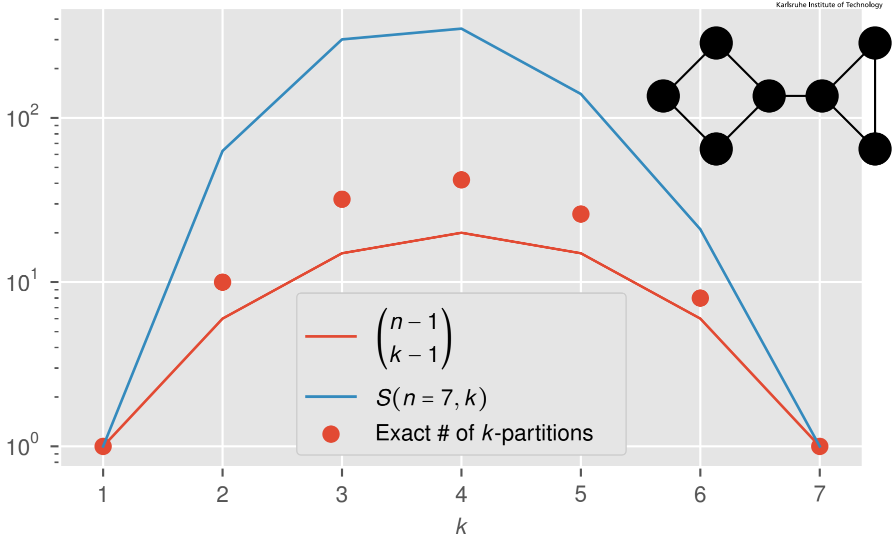
4 Small Graphs – The Graph H



4 Small Graphs – The Graph $\overline{H}$



4 Small Graphs – The Graph X_7



5 Estimated Number of k-Partitions I

Remark: Everything that follows is work in progress...

General Ideas

1. Modify the recursive formula for the Stirling numbers (2nd kind), so that it takes adjacency information into account
2. Weight the number of k-partitions by the graph density
3. Estimate the number of k-partitions as a mixture of lower and upper bound

Additional Observations:

- Number of ways to partition a graph into $n-1$ cells is $m \leq \binom{n}{2}$
- If the graph is complete, then $m = \binom{n}{2} = S(n, n-1)$

5 Estimated Number of k-Partitions I

Karlsruhe Institute of Technology

Remark: Everything that follows is work in progress…

General Ideas

1. Modify the recursive formula for the Stirling numbers (2nd kind), so that it takes adjacency information into account
2. Weight the number of k-partitions by the graph density
3. Estimate the number of k-partitions as a mixture of lower and upper bound

Additional Observations:

- Number of ways to partition a graph into $n-1$ cells is $m \leq \binom{n}{2}$
- If the graph is complete, then $m = \binom{n}{2} = S(n, n-1)$

5 Estimated Number of k-Partitions I

Remark: Everything that follows is work in progress...

General Ideas

1. Modify the recursive formula for the Stirling numbers (2nd kind), so that it takes adjacency information into account
2. Weight the number of k-partitions by the graph density
3. Estimate the number of k-partitions as a mixture of lower and upper bound

Additional Observations:

- Number of ways to partition a graph into $n-1$ cells is $m \leq \binom{n}{2}$
- If the graph is complete, then $m = \binom{n}{2} = S(n, n-1)$

Modify the Formula

Reminder: $S(n,k) = k \cdot S(n-1,k) + S(n-1,k-1)$

- The idea stays the same, but the adjacency structure restricts k
- $S_G(n,k;m) = \overline{d}(n,k,m) \cdot S_G(n-1,k;m) + S_G(n-1,k-1;m)$
- $\overline{d}(n,k,m) \leq k$ is the expected average number of neighbors (the degree) of a node
- Normally, especially in real-world graphs, the average number of neighbors is much smaller than k

Issues

- How can $\overline{d}$ be estimated? (mean number of neighbors is $\frac{2m}{n}$)
- How does m change subject to n?
- The recursive idea of "adding" an additional element does not address *where* to add a node (and it depends!)

5 Estimated Number of k-Partitions II

Modify the Formula

Reminder: $S(n,k) = k \cdot S(n-1,k) + S(n-1,k-1)$

- The idea stays the same, but the adjacency structure restricts k
- $S_G(n,k;m) = \overline{d}(n,k,m) \cdot S_G(n-1,k;m) + S_G(n-1,k-1;m)$
- $\overline{d}(n,k,m) \leq k$ is the expected average number of neighbors (the degree) of a node
- Normally, especially in real-world graphs, the average number of neighbors is much smaller than k

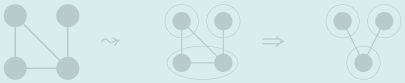
Issues

- How can $\overline{d}$ be estimated? (mean number of neighbors is $\frac{2m}{n}$)
- How does m change subject to n?
- The recursive idea of “adding” an additional element does not address *where* to add a node (and it depends!)

5 Estimated Number of k-Partitions III

Weight by Density

- We observed a decreased actual number of partitions if less edges exist
- The density $\rho = m/\binom{n}{2}$ is a measure of how many edges exist in contrast to a fully connected graph
- The actual number of graph k-partitions could be estimated as $\rho(G_{n_i=k}) \cdot S(n,k)$
- $G_{n_i=k}$ is a coarsened graph that results from the formation of clusters:



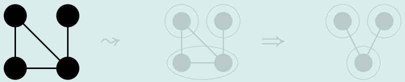
$\rho = \frac{4}{6}$ $\qquad$ $\rho(G_{n_i=3}) = \frac{2}{3}$

Issues

- As before: How to estimate the number of edges with decreasing n?

Weight by Density

- We observed a decreased actual number of partitions if less edges exist
- The density $\rho = m/\binom{n}{2}$ is a measure of how many edges exist in contrast to a fully connected graph
- The actual number of graph k-partitions could be estimated as $\rho(G_{n_i=k}) \cdot S(n,k)$
- $G_{n_i=k}$ is a coarsened graph that results from the formation of clusters:



$\rho = \frac{4}{6}$ $\qquad$ $\rho(G_{n_i=3}) = \frac{2}{3}$

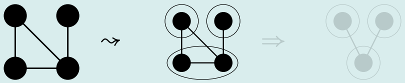
Issues

- As before: How to estimate the number of edges with decreasing n?

5 Estimated Number of k-Partitions III

Weight by Density

- We observed a decreased actual number of partitions if less edges exist
- The density $\rho = m/\binom{n}{2}$ is a measure of how many edges exist in contrast to a fully connected graph
- The actual number of graph k-partitions could be estimated as $\rho(G_{n_i=k}) \cdot S(n,k)$
- $G_{n_i=k}$ is a coarsened graph that results from the formation of clusters:



$\rho = \frac{4}{6}$ $\qquad$ $\rho(G_{n_i=3}) = \frac{2}{3}$

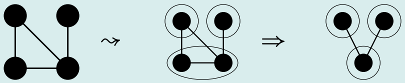
Issues

- As before: How to estimate the number of edges with decreasing n?

5 Estimated Number of k-Partitions III

Weight by Density

- We observed a decreased actual number of partitions if less edges exist
- The density $\rho = m/\binom{n}{2}$ is a measure of how many edges exist in contrast to a fully connected graph
- The actual number of graph k-partitions could be estimated as $\rho(G_{n_i=k}) \cdot S(n,k)$
- $G_{n_i=k}$ is a coarsened graph that results from the formation of clusters:



$\rho = \frac{4}{6}$ $\qquad$ $\rho(G_{n_i=3}) = \frac{2}{3}$

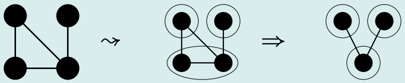
Issues

- As before: How to estimate the number of edges with decreasing n?

5 Estimated Number of k-Partitions III

Weight by Density

- We observed a decreased actual number of partitions if less edges exist
- The density $\rho = m/\binom{n}{2}$ is a measure of how many edges exist in contrast to a fully connected graph
- The actual number of graph k-partitions could be estimated as $\rho(G_{n_i=k}) \cdot S(n,k)$
- $G_{n_i=k}$ is a coarsened graph that results from the formation of clusters:



$\rho = \frac{4}{6}$ $\qquad\qquad$ $\rho(G_{n_i=3}) = \frac{2}{3}$

Issues

- As before: How to estimate the number of edges with decreasing n?

5 Estimated Number of k-Partitions IV

Estimate Mixture of Lower / Upper Bound

- The absolute lower (as presented earlier) and upper (2nd kind Stirling numbers) bounds are known
- The true number of k-partitions lies between these bounds
- This true number can be written as the linear combination $\#\text{Partitions}(n,k) = \lambda \cdot \text{lb}(n,k) + (1-\lambda)\cdot \text{ub}(n,k) = \lambda \cdot \binom{n-1}{k-1} + (1-\lambda)\cdot S(n,k)$ for some parameter $\lambda \in [0,1]$
- The estimator for λ can possibly be based on several observable graph parameters (e. g. node degree distribution)

Issues

- How to estimate λ?
- The obvious values n and m are not sufficient, as there can exist various graphs with the same n and m but a different number of partitions

5 Estimated Number of k-Partitions IV

Estimate Mixture of Lower / Upper Bound

- The absolute lower (as presented earlier) and upper (2nd kind Stirling numbers) bounds are known
- The true number of k-partitions lies between these bounds
- This true number can be written as the linear combination $\#\text{Partitions}(n,k) = \lambda \cdot \text{lb}(n,k) + (1-\lambda) \cdot \text{ub}(n,k) = \lambda \cdot \binom{n-1}{k-1} + (1-\lambda) \cdot S(n,k)$ for some parameter $\lambda \in [0,1]$
- The estimator for λ can possibly be based on several observable graph parameters (e. g. node degree distribution)

Issues

- How to estimate λ?
- The obvious values n and m are not sufficient, as there can exist various graphs with the same n and m but a different number of partitions

5 Estimated Number of k-Partitions IV

Estimate Mixture of Lower / Upper Bound

- The absolute lower (as presented earlier) and upper (2nd kind Stirling numbers) bounds are known
- The true number of k-partitions lies between these bounds
- This true number can be written as the linear combination $\#\text{Partitions}(n,k) = \lambda \cdot \text{lb}(n,k) + (1-\lambda) \cdot \text{ub}(n,k) = \lambda \cdot \binom{n-1}{k-1} + (1-\lambda) \cdot S(n,k)$ for some parameter $\lambda \in [0,1]$
- The estimator for λ can possibly be based on several observable graph parameters (e. g. node degree distribution)

Issues

- How to estimate λ?
- The obvious values n and m are not sufficient, as there can exist various graphs with the same n and m but a different number of partitions

6 Conclusion I

What is the essence of this talk?

- Clustering/Partitioning of a graph (w.r.t. the assumptions) did not get any easier
- However, the number of edges in comparison to the number of nodes (c.f. density) has a large impact on the *actual* number of possible partitions (see also Good et al. (2010), who argue on the performance of modularity clustering)
- The theoretical results and insights could help for the analysis of clustering strategies and algorithms

6 Conclusion II

What's next?

- The shown results are for graphs, which have explicit relations
- Could be applied for clustering in metric spaces as well if certain assumptions/restrictions are introduced (e. g. maximal allowed distance between the nearest data points in a cluster)
- Not mentioned here: (Graph) Symmetry restricts the search space as well (Ball, 2019)

Conjecture

As large real-world graphs are generally very sparse, the actual number of clustering partitions (the size of the search space) asymptotically approaches the lower bound.

What's next?

- The shown results are for graphs, which have explicit relations
- Could be applied for clustering in metric spaces as well if certain assumptions/restrictions are introduced (e. g. maximal allowed distance between the nearest data points in a cluster)
- Not mentioned here: (Graph) Symmetry restricts the search space as well (Ball, 2019)

Conjecture

As large real-world graphs are generally very sparse, the actual number of clustering partitions (the size of the search space) asymptotically approaches the lower bound.

Thank you for your attention!

Find the slides of this talk and additional material on Github:
`https://github.com/KIT-IISM-EM/ECDA2019`

References I

F. Ball. „Impact of Symmetries in Graph Clustering“. PhD thesis. Karlsruhe: Karlsruhe Institute of Technology, 2019. 238 pp. DOI: 10.5445/IR/1000090492.

E. T. Bell. „Exponential Polynomials“. In: *Annals of Mathematics* 35.2 (1934), pp. 258–277. ISSN: 0003-486X. DOI: 10.2307/1968431.

W. E. Donath and A. J. Hoffman. „Lower Bounds for the Partitioning of Graphs“. In: *IBM Journal of Research and Development* 17.5 (1973), pp. 420–425. ISSN: 0018-8646. DOI: 10.1147/rd.175.0420.

B. H. Good, Y.-A. de Montjoye, and A. Clauset. „Performance of Modularity Maximization in Practical Contexts“. In: *Physical Review E* 81.4 (2010), p. 046106. DOI: 10.1103/PhysRevE.81.046106.

References II

B. C. Rennie and A. J. Dobson. „On Stirling Numbers of the Second Kind". In: *Journal of Combinatorial Theory* 7.2 (1969), pp. 116–121. ISSN: 0021-9800. DOI: 10.1016/S0021-9800(69)80045-1.

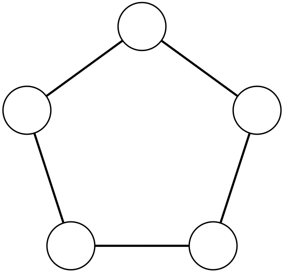
8 2-Regular Graphs I

- We have proven an exact bound for trees
- Another quite restricted class are regular graphs:
  - The degree of a node is the number of neighbors
  - In a k-regular graph, every node has exactly k neighbors
  - Special case: 2-regular graphs are determined only by the number of nodes $n \geq 3$

The 2-regular graph with $n = 5$ nodes

8 2-Regular Graphs I

- We have proven an exact bound for trees
- Another quite restricted class are regular graphs:
  - The degree of a node is the number of neighbors
  - In a k-regular graph, every node has exactly k neighbors
  - Special case: 2-regular graphs are determined only by the number of nodes $n \geq 3$



The 2-regular graph with $n = 5$ nodes

Theorem (Number of k-Partitions of a 2-regular graph)

The number of partitions of a 2-regular graph into k parts is $\binom{n}{n-k}$ for $k > 1$. For $k = 1$ it is 1.

Proof.

- For $k = 1$ there clearly exists only the trivial partition that consists of one large cell containing all nodes.
- For $1 < k \leq n$ choosing $n - k$ out of n edges can be thought of removing the edge and merging the two incident nodes:
  - Every merger means to merge two cells
  - The result is a coarsened 2-regular graph of k nodes
  - Each node represents a cell of the partition
  - How many ways exist to pick $n - k$ edges?

  $\Rightarrow$ It is just $\binom{n}{n-k}$

$\square$

Theorem (Number of k-Partitions of a 2-regular graph)

The number of partitions of a 2-regular graph into k parts is $\binom{n}{n-k}$ for $k > 1$. For $k = 1$ it is 1.

Proof.

- For $k = 1$ there clearly exists only the trivial partition that consists of one large cell containing all nodes.
- For $1 < k \leq n$ choosing $n - k$ out of n edges can be thought of removing the edge and merging the two incident nodes:
  - Every merger means to merge two cells
  - The result is a coarsened 2-regular graph of k nodes
  - Each node represents a cell of the partition
  - How many ways exist to pick $n - k$ edges?

  $\Rightarrow$ It is just $\binom{n}{n-k}$

$\square$

Hint

- The binomial coefficient has the property $\binom{n}{n-k} = \binom{n}{k}$
- Choosing $\binom{n}{k}$ edges means to separate the 2-regular graph into k cells (for $k > 1$; c.f. our argument for partitioning trees)
- To actually separate the graph into two disconnected subgraphs, we must remove two edges
- This is why picking $\binom{n}{n-1} = \binom{n}{1} = n$ is not the correct solution of the number of partitions into $k = 1$ cells

8 2-Regular Graphs III

Hint

- The binomial coefficient has the property $\binom{n}{n-k} = \binom{n}{k}$
- Choosing $\binom{n}{k}$ edges means to separate the 2-regular graph into k cells (for $k > 1$; c.f. our argument for partitioning trees)
- To actually separate the graph into two disconnected subgraphs, we must remove two edges
- This is why picking $\binom{n}{n-1} = \binom{n}{1} = n$ is not the correct solution of the number of partitions into $k = 1$ cells

8 2-Regular Graphs IV

Theorem (Number of Partitions of a 2-regular graph)

The number of partitions of a 2-regular graph is $2^n - n$.

Proof.

- We use the binomial theorem again: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$ for $x = y = 1$
- $2^n = \sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{n-k} = \underbrace{\binom{n}{n}}_{=1} + \underbrace{\binom{n}{n-1}}_{=n} + \underbrace{\sum_{k=2}^{n} \binom{n}{n-k}}_{\text{divide into } k=2,\ldots,n \text{ cells}}$
- The first term represents the number of partitions into $k = 1$ cells
- The second term is "too much"

$\Rightarrow$ $2^n - \binom{n}{n-1} = 2^n - n$

□

8 2-Regular Graphs IV

Theorem (Number of Partitions of a 2-regular graph)

The number of partitions of a 2-regular graph is $2^n - n$.

Proof.

- We use the binomial theorem again: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$ for $x = y = 1$
- $2^n = \sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{n-k} = \underbrace{\binom{n}{n}}_{=1} + \underbrace{\binom{n}{n-1}}_{=n} + \underbrace{\sum_{k=2}^{n} \binom{n}{n-k}}_{\text{divide into } k=2,\ldots,n \text{ cells}}$
- The first term represents the number of partitions into $k = 1$ cells
- The second term is “too much”

$\Rightarrow$ $2^n - \binom{n}{n-1} = 2^n - n$

□

Karlsruhe Institute of Technology

Theorem (Number of Partitions of a 2-regular graph)

The number of partitions of a 2-regular graph is $2^n - n$.

Proof.

- We use the binomial theorem again: $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^{n-k} y^k$ for $x = y = 1$
- $2^n = \sum_{k=0}^{n} \binom{n}{k} = \sum_{k=0}^{n} \binom{n}{n-k} = \underbrace{\binom{n}{n}}_{=1} + \underbrace{\binom{n}{n-1}}_{=n} + \underbrace{\sum_{k=2}^{n} \binom{n}{n-k}}_{\text{divide into } k=2,\ldots,n \text{ cells}}$
- The first term represents the number of partitions into $k = 1$ cells
- The second term is “too much”

$\Rightarrow 2^n - \binom{n}{n-1} = 2^n - n$

$\square$