

Predicting Functional Groups of Genes Through Dynamic Community Detection in Gene Co-expression Networks of *Saccharomyces cerevisiae*

CS 198 Proposal: Atienza & Maximo
Adviser: Clemente

01

Introduction

Overview of Biological Networks



Biological systems are inherently complex.

They involve numerous interacting components, such as genes, proteins, and metabolites, all working together to regulate life processes (Barabási and Oltvai, 2004).



They exhibit community structures.

These may pertain to functional modules which work to perform a biological role. It is an interesting area of research for analyzing how certain molecules are grouped within cells (Yu et al., 2013).

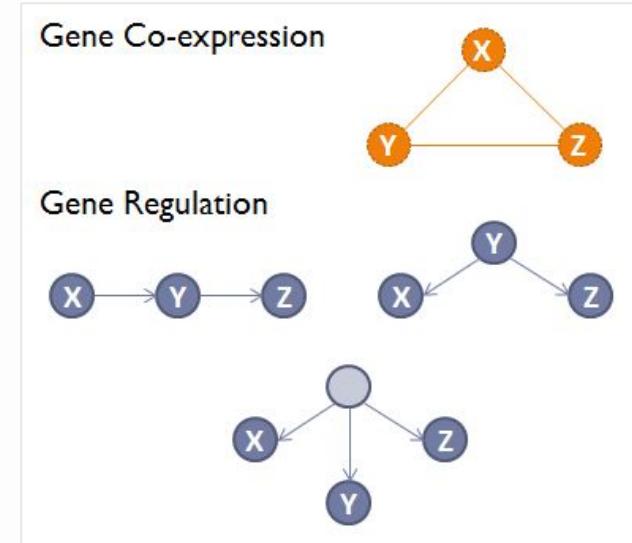


They are not static.

The dynamic nature is evident in gene co-expression networks. Studying the structure within these networks can provide insights on the regulation of biological functions and the interactions between genes.

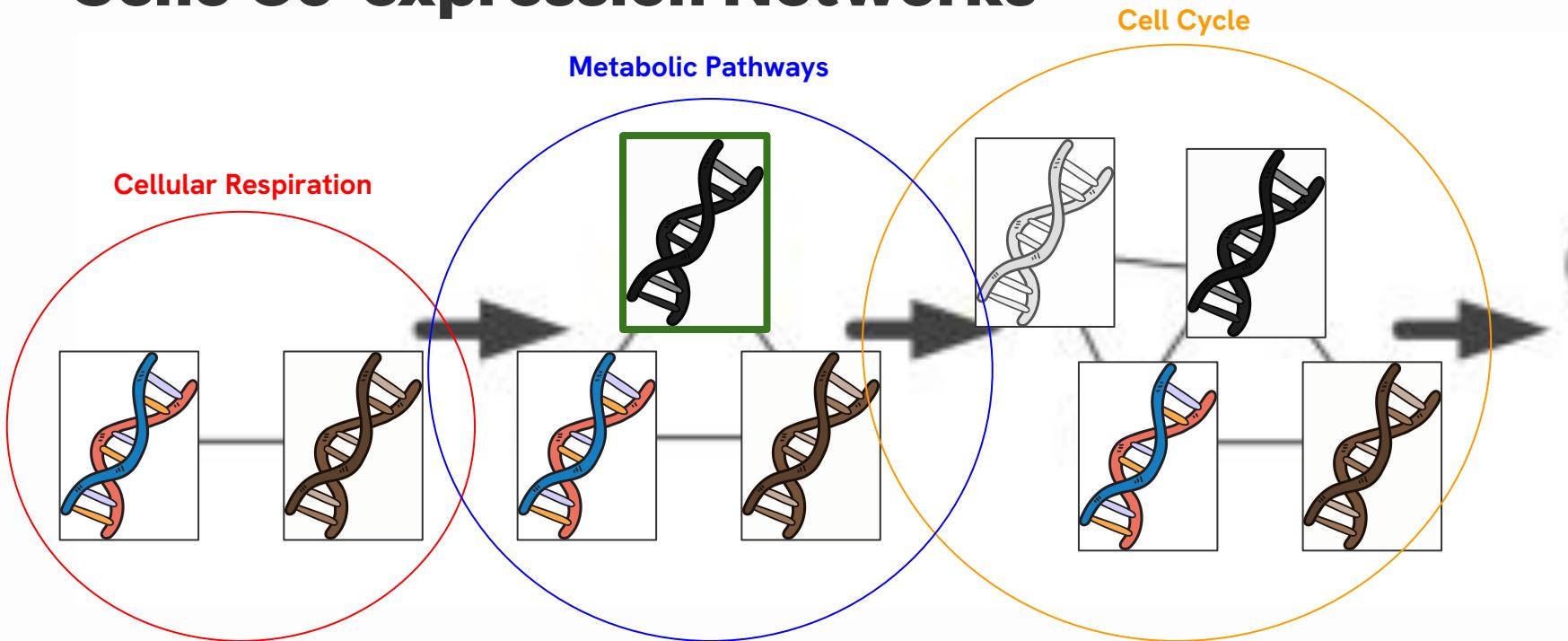
Gene Co-expression Networks

- Biological Network
- Nodes - Genes; Edges - Coexpression Relationship
 - A measure of similar patterns of activity (i.e. Expression Levels) across different conditions between genes.
 - If a pair of genes have their activity go up or down together, they are said to be coexpressed.
 - They must work together! (Stuart et al, 2003; Owens et al., 2021)



https://en.wikipedia.org/wiki/Gene_co-expression_network

Gene Co-expression Networks



We Want to Detect Communities in Gene Networks!

Gene Co-expression Networks

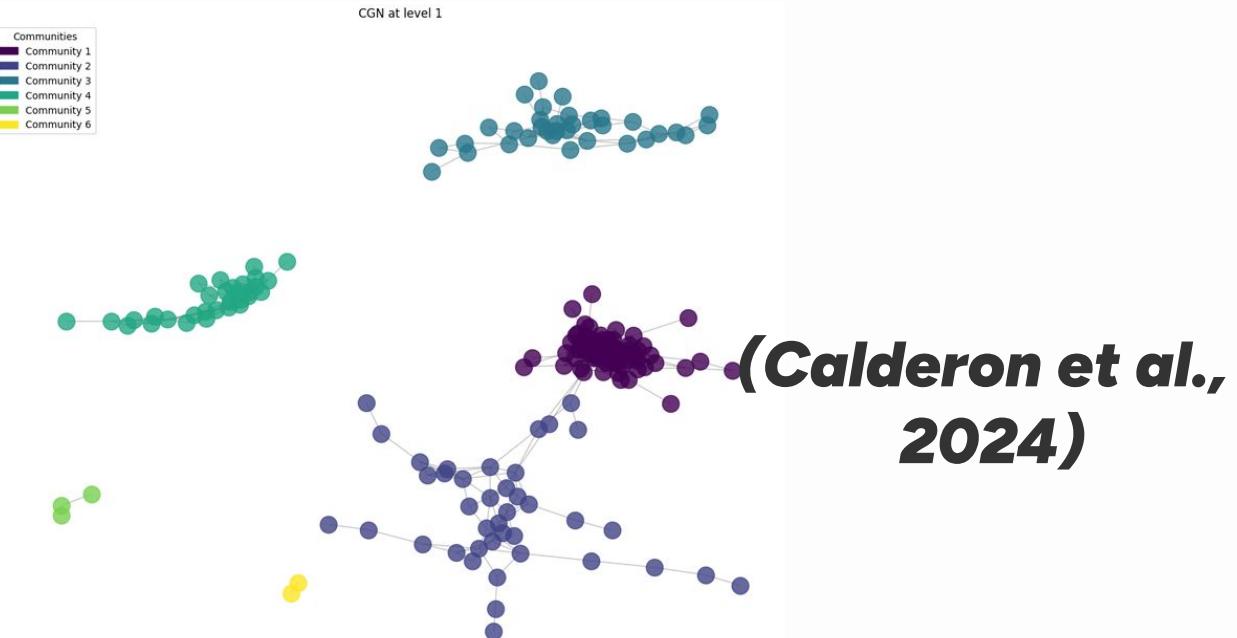


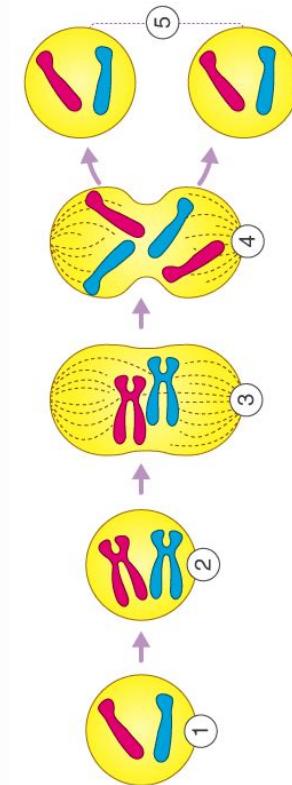
Fig. 6. Communities Detected Using GN at Level 1

Review of Related Literature

- Time-Series Gene Expression Data of *S. cerevisiae*
- Hierarchical Community Detection and the ARI
Agreement Measure in Gene Co-expression Networks
- Other DCD Algorithms
- Evaluation metrics (internal and external)

Saccharomyces cerevisiae

- Well documented yeast cells, 6000 genes
- Dataset from Cho et al (1998)
 - Genes across the different phases of the mitotic cell cycle
 - Time-series data of gene expression
 - How do the gene networks change at different time points?

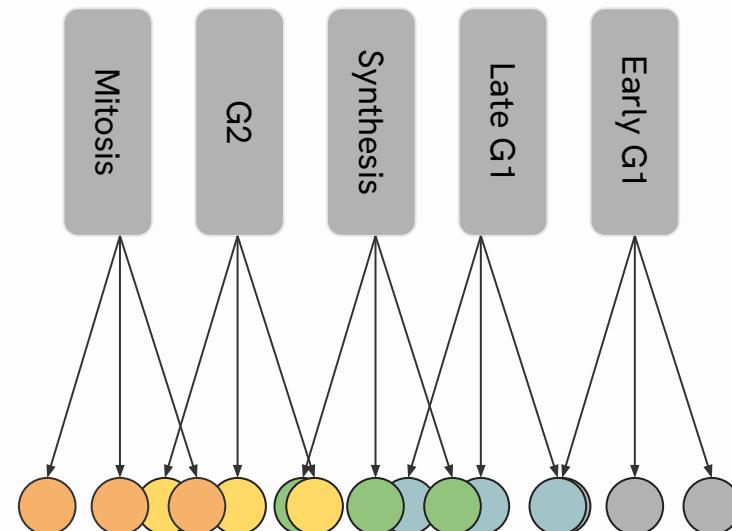
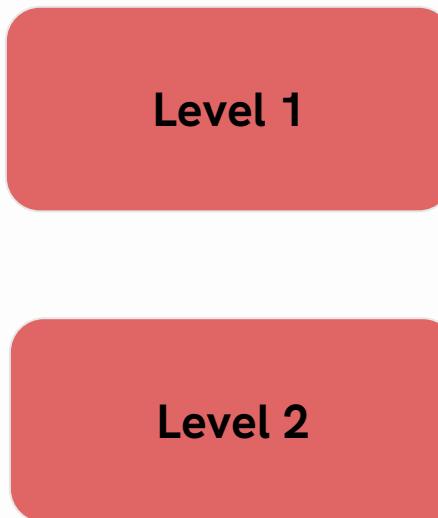


1 Interphase | 2 Prophase | 3 Metaphase | 4 Anaphase | 5 Telophase

Time-Series Gene Expression Data of *S. cerevisiae*

- Cho et al. (1998)
 - *S. cerevisiae* across different phases of the mitotic cell cycle.
 - 416 genes, **17 different time points**
 - This shows its dynamic nature!
 - Classify genes according to activity:
 1. Early G1 Phase
 2. Late G1 Phase
 3. Synthesis Phase
 4. G2 Phase
 5. Mitosis
 6. Multi-phase

Time-Series Gene Expression Data of S. *Cho Cerevisiae Dataset Structure*



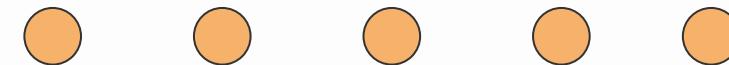
Time-Series Gene Expression Data of *S. cerevisiae*

- Yeung et al. (2001)
 - Reduced down to 384 genes, still 17 different time points
 - Removed Multi-phase
- Calderon et al. (2024)
 - Cho dataset (with Yeung's revisions), reconcile with the more modern Kyoto Encyclopedia of Genes and Genomes (KEGG).
 - Check agreement between Cho (2 Levels) and KEGG (3 Levels)

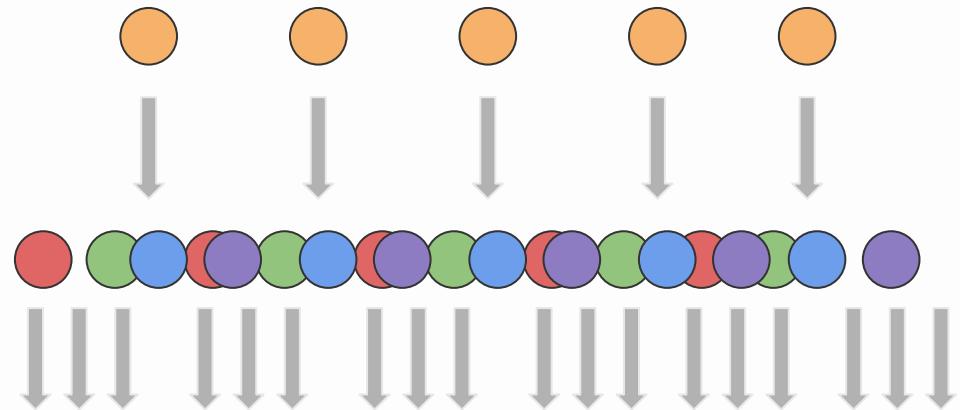
Time-Series Gene Expression Data of S. cerevisiae

KEGG Dataset Structure

Level 1



Level 2

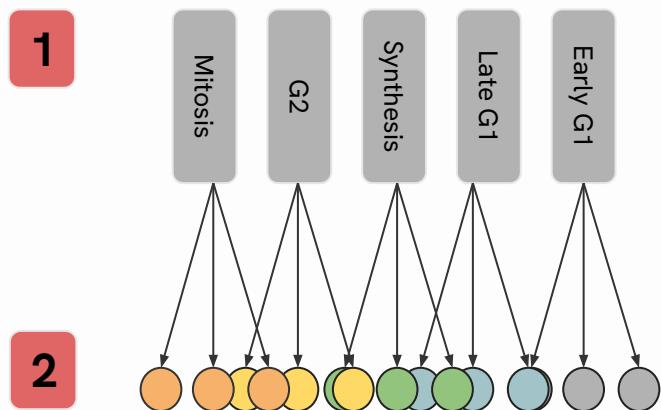


Level 3

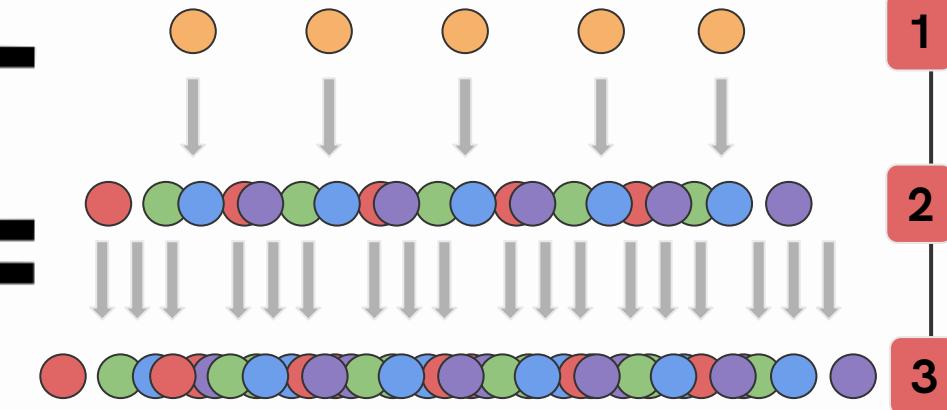


Time-Series Gene Expression Data of S. cerevisiae

Cho Dataset Structure



KEGG Dataset Structure



Time-Series Gene Expression Data of S. cerevisiae

- Calderon et al. (2024)
 - Use Adjusted Rand Index (ARI) to see how well Cho and KEGG “agree”

KEGG Level	CHO Level	ARI
1	1	0.022699
1	2	0.083272
2	1	0.057357
2	2	0.118445
3	1	0.043423
3	2	0.156492

ARI should be at least 0.5 to be relevant!

Community Detection Algorithms

- Formal Definition
 - Utilize an objective function that is able to capture the essence of a community as a group of nodes with better internal connections than external
 - Find nodes that approximate the objective function and can identify as communities by incorporating various heuristics and/or approximation algorithms

Community Detection Algorithms

- Single-objective optimization problem: identify the partition C for which

$$P(C^*) = \min_{C \in \Omega} P(C)$$

- Ω : Set of possible partitions
- C: Community structure of network G
- $P : \Omega \rightarrow \mathbb{R}$ (measure function - assume minimized to avoid loss of generality)
- Most CD algorithms are derived from the optimization problem above.

Community Detection Algorithms

- NP-hard
 - **Modularity maximization** is a critical factor that contributes to the computational complexity of the community detection problem.
 - Evaluate quality of community structure
 - Proven to be **NP-hard** (Girvan and Newman, 2002; Estivill-Castro and Parsa, 2016). **Computationally intractable.**

Community Detection Algorithms

- Reduction from known NP-hard problems
 - Estivill-Castro and Parsa (2016) showed the NP-hardness of the detection of uniform connected k -communities.
 - Stricter partitioning of graph into k communities of equal size
 - Any two nodes within a community are connected
 - Their proof involved reducing the 3-partition problem, a known NP-complete problem, into the aforementioned community detection problem.

Community Detection Algorithms

- Heuristics and Approximation
 - Girvan and Newman (2002) pioneered a groundbreaking approach focused on the concept of edge betweenness to address the computational complexity of the CD problem.
 - Iteratively removing edges that are between communities rather than within them
 - Edge betweenness → Measure the number of shortest paths between pairs of nodes that pass through a particular edge
 - Computationally expensive; worst case is $O(m^2n)$, where m is the number of edges, and n is the number of nodes

Community Detection Algorithms

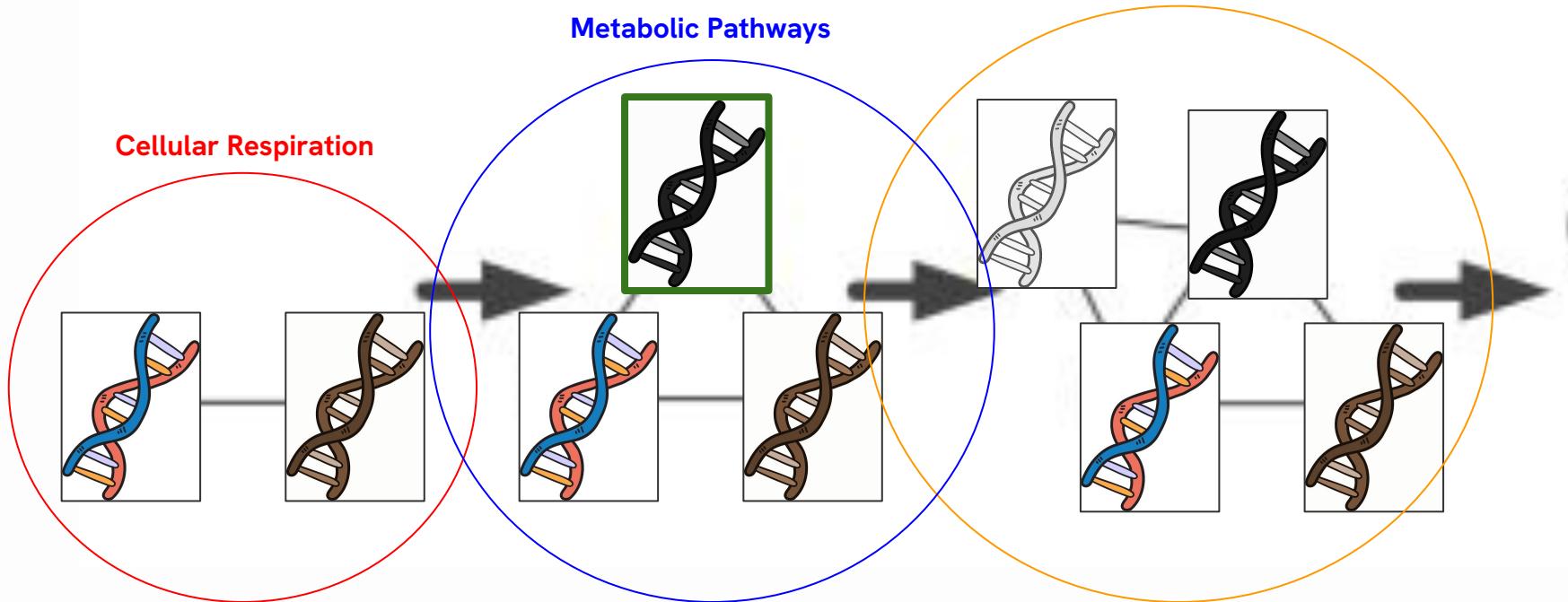
- Two categorizations of CD algorithms
 - Static
 - Used for datasets/networks that are fixed and unchanging across different time points
 - Dynamic
 - Used for datasets/networks that evolve over time to track their evolution over time

Hierarchical Community Detection and ARI in GCN

- Calderon et al. (2024)
 - Use hierarchical community detection (HCD) in GCN of *S. cerevisiae*
 - Identify functional gene communities
 - Use three HCDs:
 - Girvan-Newman
 - Paris
 - Local Optimization Function Model

The Essence of Detecting Communities...

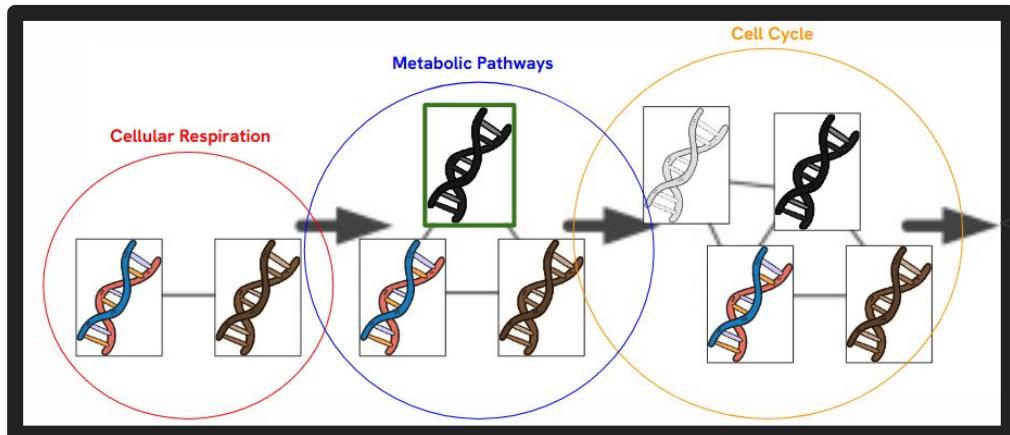
Cell Cycle



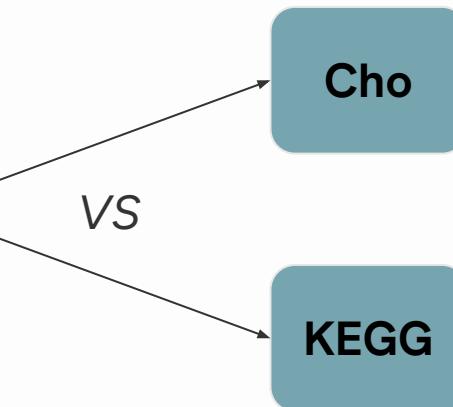
These Communities Mean a Biological Function!

How To Test the Accuracy?

“Predicted” Communities



Actual Functional Groupings



a.k.a Detected Functional Groupings

Hierarchical Community Detection and ARI in GCN

- Calderon et al. (2024)
 - Use ARI (again!) to evaluate the agreement of detected gene communities (formed by HCD) with Cho and KEGG data.
 - Result: Relatively Low ARI scores across all HCD algorithms, on both datasets
 - Implication: Either HCD or ARI cannot fully capture the GCN of *S. cerevisiae*!
 - We want to try and improve this with DCD, and a metric other than ARI.

Dynamic Community Detection Algorithms

Algorithm	Key Features	Advantages
TILES	<ul style="list-style-type: none">- Detects and tracks overlapping communities- “Domino tile” approach- Two memberships:<ul style="list-style-type: none">- <i>Core</i>: Part of a triangle within the same community- <i>Peripheral</i>: Connected to a core node but not part of a triangle- Core nodes propagate membership	<ul style="list-style-type: none">- Online nature:<ul style="list-style-type: none">- Localized updates- Parallel execution- Faster execution- Accurate for ground-truth communities- Captures evolutionary behaviors
InfoMap	<ul style="list-style-type: none">- Community detection via map equation- Based on Minimum Description Length (MDL)- Groups nodes to minimize the map equation- Iterative process: modules are treated as nodes for refinement- Utilizes flows	<ul style="list-style-type: none">- Robust for graphs with good internal flows

03

Statement of the Problem and Objectives

Statement of the Problem

Kegg and Cho

KEGG Level	CHO Level	ARI
1	1	0.022699
1	2	0.083272
2	1	0.057357
2	2	0.118445
3	1	0.043423
3	2	0.156492

HCD and Cho

ARI (Level 2)	ARI (Level 3)
0.063101	0.036326
0.061727	0.036023
0.065596	0.035593
0.081454	0.055733
0.080478	0.056259
0.08014	0.052442
0.080189	0.04908497428
0.078276	0.046109

- Calderon et. al (2024) used HCD and ARI to evaluate the detected gene communities and known biological classifications.
- However, their analysis yielded weak agreement scores.
 - Hierarchical approach may not capture the full structure
 - OR there is a better metric other than ARI

Statement of the Problem

*How can the detection of gene communities
in *S. cerevisiae* be improved
to achieve stronger agreement?*

Objectives of the Study

1. Evaluate alternative similarity metrics for improved community consistency.
2. Predict functional classifications.
3. Compare the performance of dynamic and hierarchical community detection algorithms.

03

Methodology

Overview of Methodology

1. Design a new metric MRI based on RI. Perform initial tests on sample partitions.
2. Construct the GCN. Perform Dynamic Community Detection with TILES and InfoMap algorithms.
3. Compare our detected communities in (2) with the work of Calderon et. al. (2024) by using our ground truth datasets (KEGG, Cho). Use the MRI, RI.

1.1) Limitations of the Rand Index (RI)

- Calderon, et. al (2024) utilized the Adjusted Rand Index (ARI) metric, which is the adjusted version of the Rand Index (RI).
- We inspect the formula of Rand Index according to Baker, et. al (2018).

$$Rand(A, B) = \frac{n_{00} + n_{11}}{n_{00} + n_{11} + n_{01} + n_{10}}$$

1.1) Limitations of the Rand Index (RI)

- Given a network G of i nodes, let A and B be two partitions of G . Then A and B contains the set of communities on the alphabet of G .
- For example,

$$G = \{a, b, c, d, e, f\}$$

$$A = \{\{a, b\}, \{c, d, e\}, \{f\}\}$$

$$B = \{\{a, b, c\}, \{d, e, f\}\}$$

$$G = \{a, b, c, d, e, f\}$$

$$A = \{\{a, b\}, \{c, d, e\}, \{f\}\}$$

$$B = \{\{a, b, c\}, \{d, e, f\}\}$$

$$Rand(A, B) = \frac{n_{00} + n_{11}}{n_{00} + n_{11} + n_{01} + n_{10}}$$

- n_{00} : node pairs in G that are in different communities in A and different communities in B. (e.g. the node pair (a,e))
- n_{11} : node pairs in G that are in the same community in A and B. (e.g. the node pair (a,b)).
- n_{01} : node pairs in G that are in different communities in A, but same community in B. (e.g. the node pair (a, c)).
- n_{10} : node pairs in G that are in the same community in A, but different communities in B. (e.g. the node pair (c, d)).
- **Rand(A,B) = sum of agreeing pairs / total number of pairs**

$$G = \{a, b, c, d, e, f\}$$

$$A = \{\{a, b\}, \{c, d, e\}, \{f\}\}$$

$$B = \{\{a, b, c\}, \{d, e, f\}\}$$

(a,b)

(a,c)

(a,d)

(a,e)

(a,f)

(b,c)

(b,d)

(b,e)

(b,f)

(c,d)

(c,e)

(c,f)

(d,e)

(d,f)

(e,f)

n_{00}	n_{11}	n_{01}	n_{10}
(a,d)	(a,b)	(a,c)	(c,d)
(a,e)	(d,e)	(b,c)	(c,e)
(a,f)		(d,f)	
(b,d)		(e,f)	
(b,e)			
(b,f)			

$$\text{Rand}(A,B) = (7 + 2)/(7+2+4+2)$$

$$= 0.6 \text{ (slightly significant result)}$$

1.1) Limitations of the Rand Index (RI)

- Recall: Given a network G of i nodes, let A and B be two partitions of G . Then A and B contains the set of communities on the nodes of G .
- However this assumes that **A and B come from a singular data source**. That is, we have the same nodes.
- Which is not the case for our study where we have two datasets: Cho and KEGG.
 - Cho contains fewer genes than KEGG. They have **different nodes**.
- For example:
 - Assume A is derived from Cho and B is derived from KEGG.

$$A = \{\{a, b, g\}, \{c, d, e\}, \{f\}\}$$

$$B = \{\{a, b, c, x\}, \{d, e, f\}\}$$

$$Rand(A, B) = \frac{n_{00} + n_{11}}{n_{00} + n_{11} + n_{01} + n_{10}}$$

- Suppose we find all possible node pairs and classify them, as we usually do in computing the Rand Index. How then do we classify the node pair (a, x) ?
 - Well, it's in a community in B...
 - But how do we do so for A? The node x is not part of the alphabet of A.

$$A = \{\{a, b, g\}, \{c, d, e\}, \{f\}\}$$

$$B = \{\{a, b, c, x\}, \{d, e, f\}\}$$

(a,b)	(c,d)
(a,c)	(c,e)
(a,d)	(c,f)
(a,e)	(c,x)
(a,f)	
(a,x)	
	(d,e)
	(d,f)
(b,c)	(d,x)
(b,d)	
(b,e)	
(b,f)	(e,f)
(b,x)	(e,x)

$$Rand(A, B) = \frac{n_{00} + n_{11}}{n_{00} + n_{11} + n_{01} + n_{10}}$$

n_{00}	n_{11}	n_{01}	n_{10}	??
(a,d)	(a,b)	(a,c)	(c,d)	(a,x)
(a,e)	(d,e)	(b,c)	(c,e)	(b,x)
(a,f)	(a,f)	(d,f)		(c,x)
(b,d)		(e,f)		(d,x)
	(b,e)			(e,x)
	(b,f)			
	(c,f)			

$$Rand(A, B) = ???$$

1.2) Designing the Modified Rand Index (MRI)

$$MRI(A, B) = \frac{n_{00} + n_{11} + n_{xx}}{n_{00} + n_{11} + n_{01} + n_{10} + n_{xx}}$$

- n_{xx} : the number of pairs where at least one node in the pair does not exist in one of the partitions, ensuring that pairs missing from either A or B are appropriately categorized.
- Test this new metric on some sample partitions

$$A = \{\{a, b, g\}, \{c, d, e\}, \{f\}\}$$

$$MRI(A, B) = \frac{n_{00} + n_{11} + n_{xx}}{n_{00} + n_{11} + n_{01} + n_{10} + n_{xx}}$$

$$B = \{\{a, b, c, x\}, \{d, e, f\}\}$$

(a,b)	(c,d)
(a,c)	(c,e)
(a,d)	(c,f)
(a,e)	(c,x)
(a,f)	
(a,x)	(d,e)
	(d,f)
(b,c)	(d,x)
(b,d)	
(b,e)	(e,f)
(b,f)	(e,x)
(b,x)	

n ₀₀	n ₁₁	n ₀₁	n ₁₀	n _{xx}
(a,d)	(a,b)	(a,c)	(c,d)	(a,x)
(a,e)	(d,e)	(b,c)	(c,e)	(b,x)
(a,f)		(d,f)		(c,x)
(b,d)		(e,f)		(d,x)
	(b,e)			(e,x)
	(b,f)			
	(c,f)			

$$\begin{aligned} \text{Rand}(A, B) &= (7 + 2 + 5)/20 \\ &= 0.7 \text{ (statistically significant)} \end{aligned}$$

1.3) Testing the Modified Rand Index (MRI)

$$MRI(A, B) = \frac{n_{00} + n_{11} + n_{xx}}{n_{00} + n_{11} + n_{01} + n_{10} + n_{xx}}$$

- **Ideal Case:** Generate two partitions with perfect agreement. MRI should yield a score of 100%.
- **General Case:** Generate two random partitions, reflecting real world scenarios.
- **Non-Ideal Case:** Generate two partitions with drastically varying levels of disagreement.

2.1) Dynamic Community Detection (Forming GCN)

- We first define the structural properties of the GCN to ensure its compatibility with DCD algorithms.
- Edge-Weighted and Undirected
 - Relationships between genes using weighted edges based on their co-expression values
- Temporal Representation
 - Modeled dynamically across different time points to capture changes in co-expression relationships over time
- Using the above properties, we form the GCNs for both the Cho and KEGG datasets through a **value-based algorithm** for network construction.

2.1) Dynamic Community Detection (Forming GCN)

• Correlation-Based Edge Weights

- Pearson Correlation Coefficient (PCC) is used to quantify pairwise gene co-expression relationships and to measure the linear relationship between gene expression profiles across time.
- Thresholding Correlation Values
 - A threshold value ($\delta = 0.85$) is applied to determine whether an edge between two genes exists. An edge is only created when it exceeds this threshold, ensuring that only strongly co-expressed genes are connected.

2.1) Dynamic Community Detection (Forming GCN)

- Forming GCNs for Cho and KEGG

- Networks are weighted and undirected to reflect mutual interactions between genes.
- Same construction method for both datasets ensures consistency, facilitating comparisons between the networks and their detected communities.

2.2) Dynamic Community Detection (Applying)

- Identify and implement DCD algorithms for *weighted and undirected* networks
- TILES
 - Algorithm that is able to detect and track evolution of overlapping communities in dynamic networks, similar to the fall of a domino tile
- InfoMap
 - Algorithm that utilizes a map equation to detect community structures in networks and employs the concept of minimum description length in optimally computing partitions

3) Performance Comparison of Hierarchical and Dynamic

- Determine which algorithms are suitable for deriving biologically significant results
- Internal Criteria (all are available from the NetworkX library)
 - Modularity
 - Quality of a network's community structure
 - High modularity indicates denser connections between vertices in the same community compared to vertices in different communities
 - Closeness Centrality
 - Transmission rate of information to a node from all other nodes
 - Accessibility and ability to quickly capture information

3) Performance Comparison of Hierarchical and Dynamic

- External Criteria (the following are available from the Scikit-learn library)
 - Rand Index (RI)
 - Similarity between two communities
 - Considers all node pairs and determines whether they belong in the same community
 - Probability that a node pair from A and B will agree

3) Performance Comparison of Hierarchical and Dynamic

- Rand Index (ARI)

$$Rand(A, B) = \frac{n_{00} + n_{11}}{n_{00} + n_{11} + n_{01} + n_{10}}$$

- n_{00} : node pairs in G that are in different communities in A and different communities in B
- n_{11} : node pairs in G that are in the same community in A and B
- n_{01} : node pairs in G that are in different communities in A, but same community in B
- n_{10} : node pairs in G that are in the same community in A, but different communities in B
- **Rand(A,B) = sum of agreeing pairs / total number of pairs**

3) Performance Comparison of Hierarchical and Dynamic

- Normalized Mutual Information (NMI)
 - Quality of clustering results compared to ground truth
 - Confusion matrix N is defined
 - Rows → ground truth class labels
 - Columns → communities generated by a CD algorithm
 - $N_{ij} \rightarrow$ number of nodes with class label i that are in community j .

3) Performance Comparison of Hierarchical and Dynamic

- Normalized Mutual Information (NMI)

$$\text{NMI}(A, C) = \frac{-2 \sum_{i=1}^{n_A} \sum_{j=1}^{n_C} N_{ij} \log \left(\frac{N_{ij} N}{N_{i\cdot} N_{\cdot j}} \right)}{\sum_{i=1}^{n_A} N_{i\cdot} \log \left(\frac{N_{i\cdot}}{N} \right) + \sum_{j=1}^{n_C} N_{\cdot j} \log \left(\frac{N_{\cdot j}}{N} \right)}$$

- A → ground truth partition; C → CD algorithm partition
- n_A → # of ground truth communities; n_C → # of communities detected by the CD algorithm
- $N_{i\cdot}$ → # of nodes with class label i ; $N_{\cdot j}$ → # of nodes in community j
- NMI is equal to 1 if the partitions are equal; 0 otherwise

3) Performance Comparison of Hierarchical and Dynamic

- External Criteria

- Modified Rand Index (MRI)
 - Addresses a significant weakness of RI by accounting for pairs that involve nodes that are missing in one of the two networks being compared
 - Might yield stronger RI scores

04

Scope and Limitations of the Study

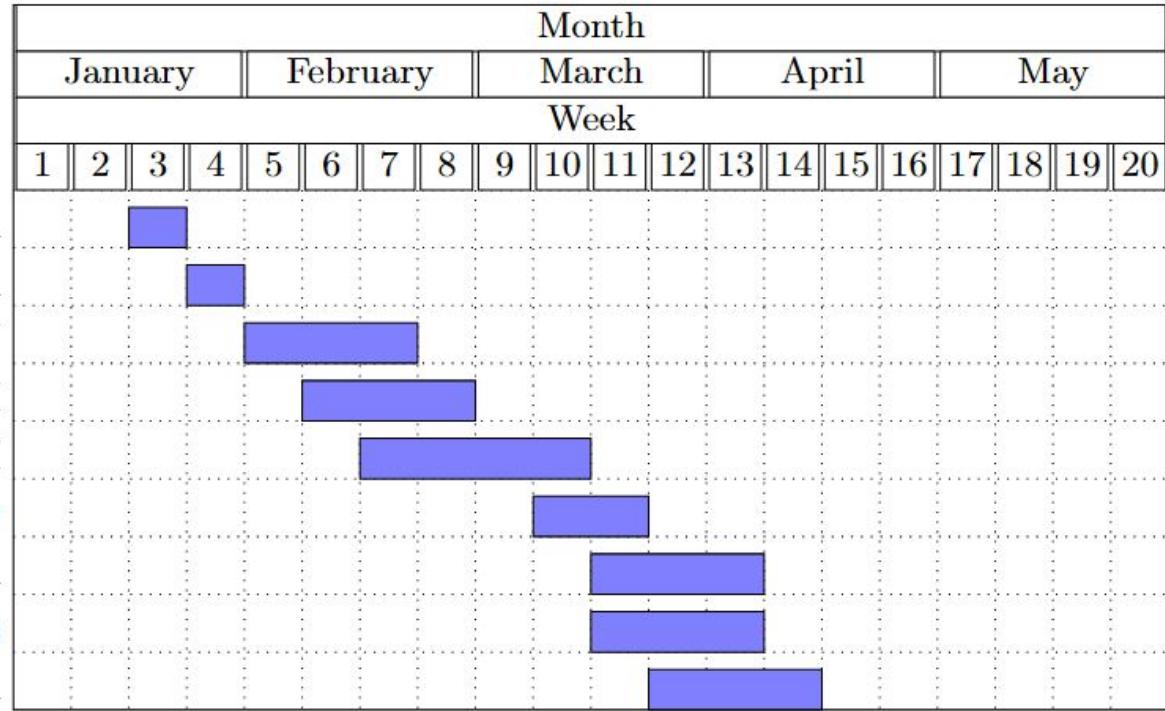
Scope and Limitations of the Study

- Primarily focused on dynamic community detection and modification of the Rand Index (RI) formula
- Same dataset as the one used by Calderon et al. (2024)
- Utilize existing Gene Co-expression Network construction algorithm (Clemente, 2022)
- Utilize only existing DCD algorithms that are proven effective in biological contexts
- Cannot directly compare our proposed metric MRI to ARI
 - Instead, MRI to RI
 - No plans of formulating an Adjusted MRI, i.e. AMRI
- Employ both existing scoring metrics and proposed metric (MRI) in assessing the performance of DCD algorithms

05

Timeline of Activities

Timeline of Activities



06

References

References

- Bakker, C., Halappanavar, M., & Visweswara Sathanur, A. (2018). Dynamic graphs, community detection, and riemannian geometry. *Applied Network Science*, 3 (1), 3. <https://doi.org/10.1007/s41109-018-0059-2>
- Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5 (2), 101-113. <https://doi.org/10.1038/nrg1272>
- Calderon, J. K., & Ventures, P. A. (2024). Hierarchical community detection on co-expression networks for functional classification of cell-cycle regulated genes of *saccharomyces cerevisiae*. Algorithms and Complexity Lab, University of the Philippines Diliman.
- Chejara, P., & Godfrey, W. W. (2017). Comparative analysis of community detection algorithms. *2017 Conference on Information and Communication Technology (CICT)*, 1-5. <https://doi.org/10.1109/INFOCOMTECH.2017.8340627>
- Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J., & Davis, R. W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Molecular Cell*, 2 (1), 65-73. [https://doi.org/10.1016/s1097-2765\(00\)80114-8](https://doi.org/10.1016/s1097-2765(00)80114-8)
- Clemente, J. B., Besas, G., Callado, J., & Evangelista, J. E. (2022). Predicting the biological classification of cell-cycle regulated genes of *saccharomyces cerevisiae* using community detection algorithms on gene co-expression networks. <https://doi.org/10.48550/ARXIV.2208.10119>
- Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances in Physics*, 51 (4), 1079-1187. <https://doi.org/10.1080/00018730110112519>

References

- Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99 (12), 7821-7826. <https://doi.org/10.1073/pnas.122653799>
- Held, P., Krause, B., & Kruse, R. (2016). Dynamic clustering in social networks using louvain and infomap method. 2016 Third European Network Intelligence Conference (ENIC), 61-68. <https://doi.org/10.1109/ENIC.2016.017>
- Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M., & Ishiguro-Watanabe, M. (2022). KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research*, 51 (D1), D587-D592. <https://doi.org/10.1093/nar/gkac963>
- Kherad, M., Dadras, M., & Mokhtari, M. (2024). Community detection based on influential nodes in dynamic networks. *The Journal of Supercomputing*, 80 (16), 24664-24688. <https://doi.org/10.1007/s11227-024-06367-4>
- Lau, L. Y., Reverter, A., Hudson, N. J., Naval-Sanchez, M., Fortes, M. R. S., & Alexandre, P. A. (2020). Dynamics of gene co-expression networks in time-series data: A case study in drosophila melanogaster embryogenesis. *Frontiers in Genetics*, 11. <https://doi.org/10.3389/fgene.2020.00517>
- Liu, X., Cheng, H.-M., & Zhang, Z.-Y. (2020). Evaluation of Community Detection Methods. *IEEE Transactions on Knowledge & Data Engineering*, 32 (09), 1736-1746. <https://doi.org/10.1109/TKDE.2019.2911943>

References

- Ma, X., & Dong, D. (2017). Evolutionary nonnegative matrix factorization algorithms for community detection in dynamic networks. *IEEE Transactions on Knowledge and Data Engineering*, 29 (5), 1045-1058. <https://doi.org/10.1109/TKDE.2017.2657752>
- Márquez, R., & Weber, R. (2023). Dynamic community detection including node attributes. *Expert Systems with Applications*, 223, 119791. <https://doi.org/https://doi.org/10.1016/j.eswa.2023.119791>
- Ovens, K., Eames, B. F., & McQuillan, I. (2021). Comparative analyses of gene co-expression networks: Implementations and applications in the study of evolution. *Frontiers in Genetics*, 12. <https://doi.org/10.3389/fgene.2021.695399>
- Parapouli, M., Vasileiadi, A., Afendra, A.-S., & Hatziloukas, E. (2020). *Saccharomyces cerevisiae* and its industrial applications. *AIMS Microbiology*, 6 (1), 1-32. <https://doi.org/10.3934/microbiol.2020001>
- Priya, A., Sharma, S., Sinha, K., & Yogesh, Y. (2023). Community detection in networks: A comparative study. *2023 International Conference on Device Intelligence, Computing and Communication Technologies*, (DICCT), 505-510. <https://doi.org/10.1109/DICCT56244.2023.10110206>
- Redekar, S. S., & Varma, S. L. (2022). A survey on community detection methods and its application in biological network, 1030-1037. <https://doi.org/10.1109/ICAAIC53929.2022.9792913>
- Rossetti, G., & Cazabet, R. (2018). Community discovery in dynamic networks: A survey. *ACM Comput. Surv.*, 51 (2). <https://doi.org/10.1145/3172867>

References

- Rossetti, G., Pappalardo, L., Pedreschi, D., & Giannotti, F. (2017). Tiles: An online algorithm for community discovery in dynamic social networks. *Machine Learning*, 106 (8), 1213-1241. <https://doi.org/10.1007/s10994-016-5582-8>
- Sattar, N. S., Buluc, A., Ibrahim, K. Z., & Arifuzzaman, S. (2023). Exploring temporal community evolution: Algorithmic approaches and parallel optimization for dynamic community detection. *Applied Network Science*, 8 (1), 64. <https://doi.org/10.1007/s41109-023-00592-1>
- Seifkar, M., Farzi, S., & Barati, M. (2020). C-blondel: An efficient louvain-based dynamic community detection algorithm. *IEEE Transactions on Computational Social Systems*, 7 (2), 308-318. <https://doi.org/10.1109/TCSS.2020.296419>
- Singh, D. K., Haraty, R. A., Debnath, N. C., & Choudhury, P. (2020). An analysis of the dynamic community detection algorithms in complex networks. 2020 IEEE International Conference on Industrial Technology (ICIT), 989-994. <https://doi.org/10.1109/ICIT45562.2020.9067224>
- Stuart, J. M., Segal, E., Koller, D., & Kim, S. K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. *Science*, 302 (5643), 249-255. <https://doi.org/10.1126/science.1087447>
- Sun, Z., Wang, B., Sheng, J., Yu, Z., Zhou, R., & Shao, J. (2019). Community detection based on information dynamics. *Neurocomputing*, 359, 341-352. <https://doi.org/https://doi.org/10.1016/j.neucom.2019.06.020>
- Yeung, K. Y., Haynor, D. R., & Ruzzo, W. L. (2001). Validating clustering for gene expression data. *Bioinformatics*, 17 (4), 309-318. <https://doi.org/10.1093/bioinformatics/17.4.309>

References

Yu, D., Kim, M., Xiao, G., & Hwang, T. H. (2013). Review of biological network data and its applications. *Genomics Inform*, 11 (4), 200-210.
<https://doi.org/10.5808/GI.2013.11.4.200>

Zhuang, D., Chang, J. M., & Li, M. (2021). Dynamo: Dynamic community detection by incrementally maximizing modularity. *IEEE Transactions on Knowledge and Data Engineering*, 33 (5), 1934-1945. <https://doi.org/10.1109/TKDE.2019.2951419>

Thanks!

Do you have any questions?

Carmelo Ellezandro Atienza | Calvin James Maximo
craatienza1@up.edu.ph | ctmaximo1@up.edu.ph

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution

Dynamic Community Detection

Algorithms

- TILES Algorithm
 - Detects and tracks the evolution of overlapping communities
 - “Domino tile” approach



Dynamic Community Detection Algorithms

- TILES Algorithm

- Two types of community memberships
 - Core
 - Part of a triangle within the same community
 - Peripheral
 - Connected to a core node but not part of any triangle within the core node's community
- Only core nodes are allowed to propagate community membership.



Dynamic Community Detection

Algorithms

- TILES Algorithm
 - Generates overlapping communities, which is identical to real-world networks
 - Advantageous online nature
 - Localized updating process which involves few nodes and communities
 - Parallel execution, which supports faster community computations, and captures evolutionary behaviors



Dynamic Community Detection Algorithms

- TILES Algorithm

- Rossetti et al. (2017) assessed TILES against various CD algorithms (DEMON, CFinder, and iLCD) on synthetic and real-world networks
 - Faster execution times
 - More accurate with ground-truth communities
 - Identified significant lifecycle events
- Showed potential in adapting to gene co-expression networks



Dynamic Community Detection

Algorithms

- InfoMap Algorithm
 - Utilizes a map equation to detect community structures
 - Community detection as a coding problem
 - Minimum Description Length (MDL)
 - Description length of a random walker is expressed by the map equation.
 - Good modular structure = small description length
 - Description length can be compressed if a random walker stays longer within a partition.

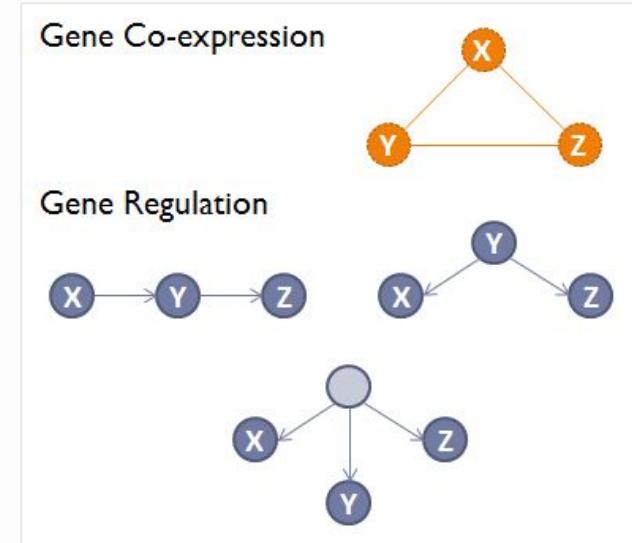
Dynamic Community Detection Algorithms

- InfoMap Algorithm

- Groups nodes through a collective process, each node a separate module
- Randomly selects nodes to combine → significant decrease in the map equation
- Previously formed modules are treated as nodes in the succeeding steps and the process is repeated until it is no longer possible to obtain a decrease in the map equation.
- Fundamental: utilization of flows
 - Robust as long as clusters flow well inside the graph
 - Fails otherwise if there is a lack of flow

Gene Co-expression Networks

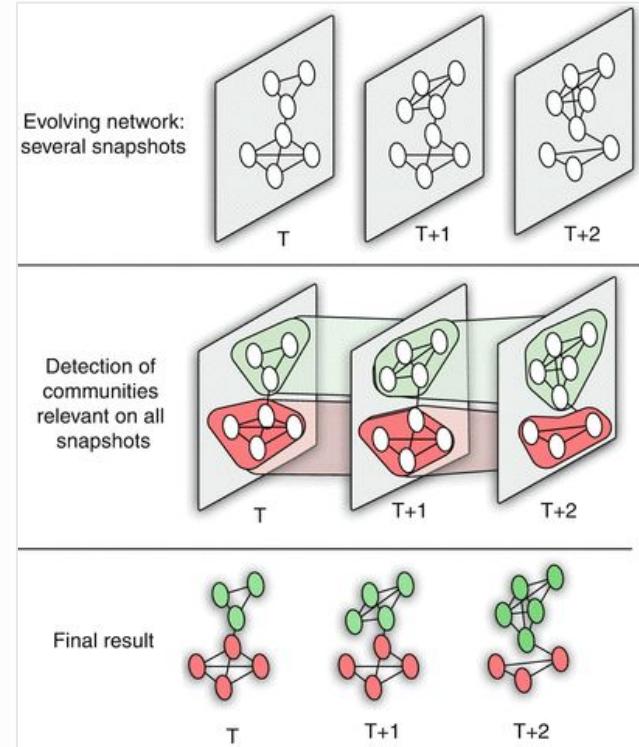
- Biological Network
- Nodes - Genes; Edges - Coexpression Relationship
 - A measure of similar patterns of activity (i.e. Expression Levels) across different conditions between genes.
 - If a pair of genes have their activity go up or down together, they are said to be coexpressed.
 - They must work together!



https://en.wikipedia.org/wiki/Gene_co-expression_network

Community Detection

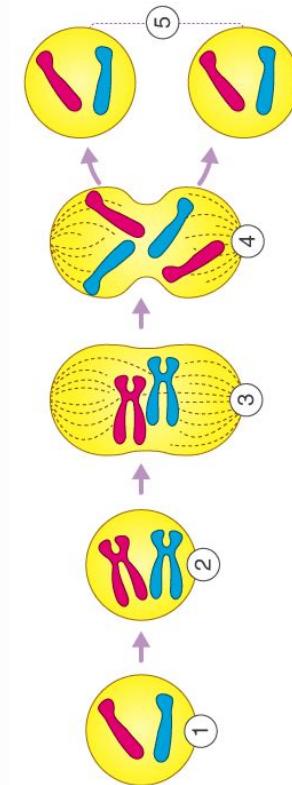
- Static Community Detection
 - Assumes the network is fixed throughout different time periods
 - Can reveal hierarchies in networks
- Dynamic Community Detection
 - Detect how communities vary at each time step
 - Can reveal temporal patterns
 - Our method of interest



https://link.springer.com/referenceworkentry/10.1007/978-1-4419-6170-8_383

Saccharomyces cerevisiae

- Well documented yeast cells, 6000 genes
- Dataset from Cho et al (1998)
 - Genes across the different phases of the mitotic cell cycle
 - Time-series data of gene expression
 - How do the gene networks change at different time points?



1 Interphase | 2 Prophase | 3 Metaphase | 4 Anaphase | 5 Telophase

Evaluating the Prediction (Internal Criteria)

- Assess clustering results without the use of external information (e.g. ground truth)
- Modularity
 - Quality of a network community structure
 - High modularity → denser intraconnections compared to interconnections
 - Evaluate the strength of algorithm in dividing a network into communities

Evaluating the Prediction (Internal Criteria)

- Closeness Centrality
 - Quantify the speed of transmission of information to a node from other nodes
 - Overall accessibility of a node and capability to obtain information rapidly

Evaluating the Prediction (External Criteria)

- Evaluate the similarity of results of algorithms against a ground truth generated by prior experiments
- Adjusted Rand Index (ARI)
 - First need to understand Rand Index (RI)
 - Pairwise comparison of nodes if they belong in the same community
 - Compare these pairs against another graph
 - Score = 1 → same community/different communities in both graphs
 - Otherwise, score = 0 → difference in community structure

Evaluating the Prediction (External Criteria)

- Adjusted Rand Index (ARI)
 - Otherwise, it gets a 0 which indicates a difference in the community structure.
 - Result of dividing the sum of scores of all node pairs and number of pairs is between 0 and 1
 - Directly proportional to similarity of two graphs' community structures

Evaluating the Prediction (External Criteria)

- Adjusted Rand Index (ARI)

$$ARI = \frac{RI - ExpectedRI}{MaxRI - ExpectedRI}$$

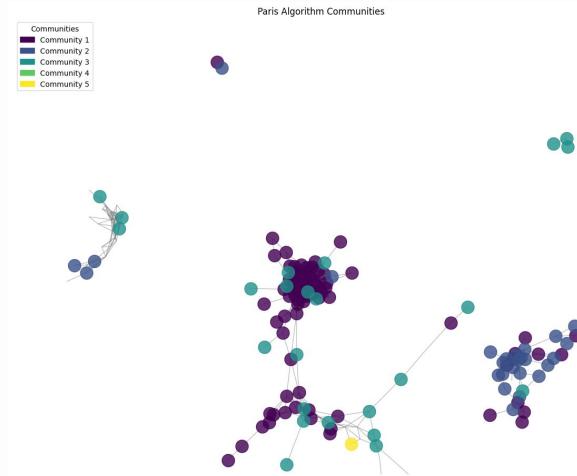
- ARI is simply an extension of RI where chance is taken into account.
- A value of 1 indicates the identicality between predicted and actual clustering.
- A negative value indicates that the cluster agreement is lower than expected by chance.

03

Related Works

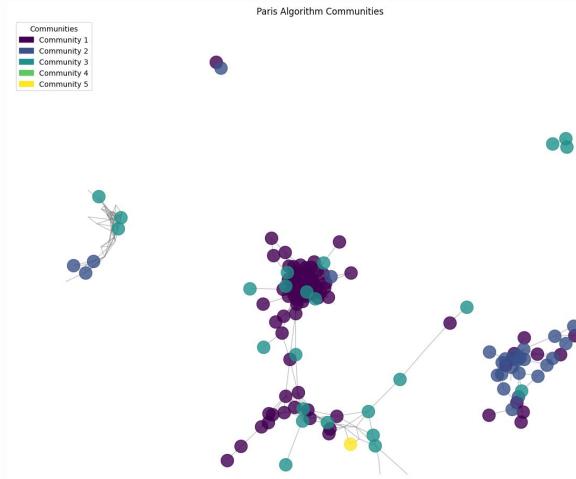
Hierarchical CD on GCNs

- Calderon and Ventures (2024)
 - Explored hierarchical CD in *S. cerevisiae* to enhance functional classification of cell-cycle regulated genes
 - Analyzed the performance of three algorithms:
 - Girvan-Newman
 - Paris
 - LFM



Hierarchical CD on GCNs

- Calderon and Ventures (2024)
 - Hierarchical CD is appropriate for GCNs due to their hierarchical nature.
 - Valuable insights into the static structure of GCNs
 - Open problem: dynamic nature of gene interactions



Dynamic Community Detection Algorithms

- C-Blondel Algorithm
 - Proposed by Seifkar et al. (2020)
 - Built upon the Louvain algorithm, and leverages historical information to improve execution time
 - Compresses the network into supernodes and superedges
 - Identifies **destructive nodes** through a degree centrality heuristic
 - Faster than other DCD algorithms (D-Blondel, S-Blondel)
 - Comparable performance in terms of modularity

Dynamic Community Detection Algorithms

- Semi-supervised ENMF (sE-NMF) Algorithm
 - Proposed by Ma and Dong (2017)
 - Combines evolutionary nonnegative matrix factorization (ENMF) and spectral clustering for dynamic community detection, balancing clustering accuracy and temporal consistency
 - Key advantage: Leveraging **a priori information**, avoiding local optima without affecting the run time complexity

Dynamic Community Detection

Algorithms

- Semi-supervised ENMF (sE-NMF) Algorithm
 - Evaluated against three well-known DCD algorithms (FacetNet, Kim-Han, DYNMOGA)
 - Particularly interesting network: breast cancer progression based on gene expression data
 - Outperformed the other algorithms in terms of specificity and sensitivity
 - Great potential as a tool for assessing temporal networks of disease progression

Thanks!

Do you have any questions?

Carmelo Ellezandro Atienza | Calvin James Maximo
craatienza1@up.edu.ph | ctmaximo1@up.edu.ph

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution

Outline

- An Overview of Biological Networks
- An Overview of Community Detection
- Gene Co-expression Networks
- Review of Related Literature
 - Hierarchical Community Detection by Ventures, Calderon
 - Dynamic Community Detection Algorithms

Contents of this template

You can delete this slide when you're done editing the presentation

<u>Fonts</u>	To view this template correctly in PowerPoint, download and install the fonts we used
<u>Used and alternative resources</u>	An assortment of graphic resources that are suitable for use in this presentation
<u>Thanks slide</u>	You must keep it so that proper credits for our design are given
<u>Colors</u>	All the colors used in this presentation
<u>Icons and infographic resources</u>	These can be used in the template, and their size and color can be edited
<u>Editable presentation theme</u>	You can edit the master slides easily. For more info, click here

For more info:

[Slidesgo](#) | [Blog](#) | [FAQs](#)

You can visit our sister projects:

[Frepik](#) | [Flaticon](#) | [Storyset](#) | [Wepik](#) | [Videvo](#)

Table of contents

01

Statement

You can describe the topic
of the section here

02

Hypothesis

You can describe the topic
of the section here

03

Objectives

You can describe the topic
of the section here

04

Methodology

You can describe the topic
of the section here

05

Analysis

You can describe the topic
of the section here

06

Conclusions

You can describe the topic
of the section here

Current situation & problems statement

Current situation

Jupiter is a gas giant and the biggest planet in the Solar System. It's the fourth-brightest object in the night sky. It was named after the Roman god of the skies and lightning



01



02



03

Earth

Earth is the third planet from the Sun and has life

Mars

Despite being red, Mars is actually a cold place

Mercury

Mercury is the smallest planet in the Solar System

Problems

Hypotheses



Hypothesis 1

Mercury is the closest planet to the Sun. The planet's name has nothing to do with the liquid metal



Hypothesis 2

Venus has a beautiful name and is the second planet from the Sun. It's hot and its atmosphere is poisonous



Hypothesis 3

Jupiter is a gas giant and the biggest planet in the Solar System. It was named after the Roman god

Study objectives



Mercury

Mercury is the closest planet to the Sun and the smallest of them all



Venus

Venus has a beautiful name and is the second planet from the Sun



Mars

Despite being red, Mars is actually a cold place. It's full of iron oxide dust

Reviewing concepts is a good idea



Mars

Mars is actually a very cold place



Venus

Venus has extremely high temperatures



Neptune

Neptune is the farthest planet from the Sun



Mercury

Mercury is the closest planet to the Sun

Saturn

Saturn is a gas giant with several rings



Jupiter

Jupiter is the biggest planet of them all

Literature review

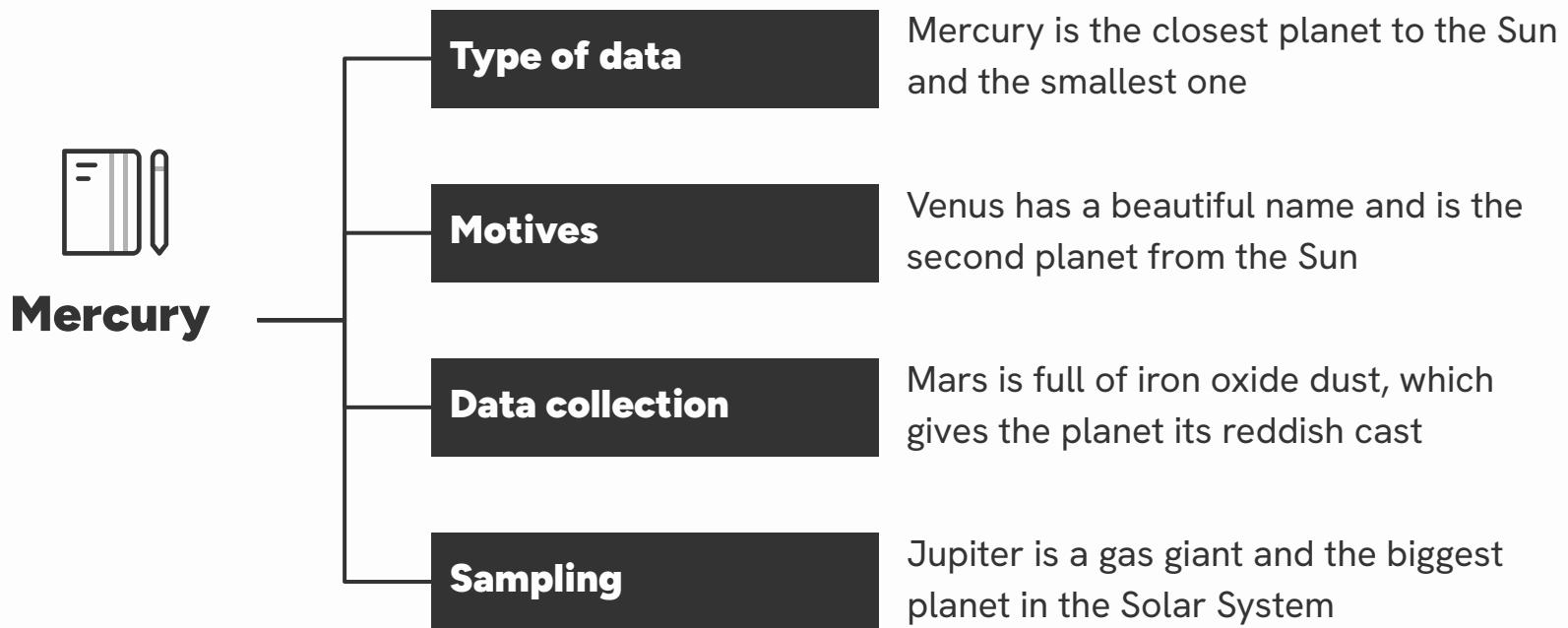
- AUTHOR. (YEAR). *Title of the publication*. Publisher
 - Mercury is the closest planet to the Sun and the smallest one
- AUTHOR. (YEAR). *Title of the publication*. Publisher
 - Mars is full of iron oxide dust, which gives the planet its reddish cast
- AUTHOR. (YEAR). *Title of the publication*. Publisher
 - Jupiter is a gas giant and the biggest planet in the Solar System
- AUTHOR. (YEAR). *Title of the publication*. Publisher
 - Venus has a beautiful name and is the second planet from the Sun
- AUTHOR. (YEAR). *Title of the publication*. Publisher
 - Earth is the third planet from the Sun and harbors life

Theoretical framework

Theoretical framework		
Key terms	Relevant theories	Our framework
<ul style="list-style-type: none">• Mercury is small• Earth harbors life• Jupiter is big	<p>Theory 1</p> <p>Saturn is a gas giant and has rings. It's composed mostly of hydrogen and helium</p> <p>Theory 2</p> <p>Neptune is the farthest planet from the Sun in Solar System and also an ice giant</p>	 <p>Venus has a beautiful name and is the second planet from the Sun. It's very hot</p>

Schedule

Methodology



Analysis & development

Phase 01

- Mercury is the closest planet to the Sun and the smallest one in the Solar System—it's a bit larger than the Moon
- Jupiter is a gas giant, the biggest planet in the Solar System and the fourth-brightest object in the night sky
- Neptune is the farthest planet from the Sun. It's also the fourth-largest planet by diameter in the Solar System

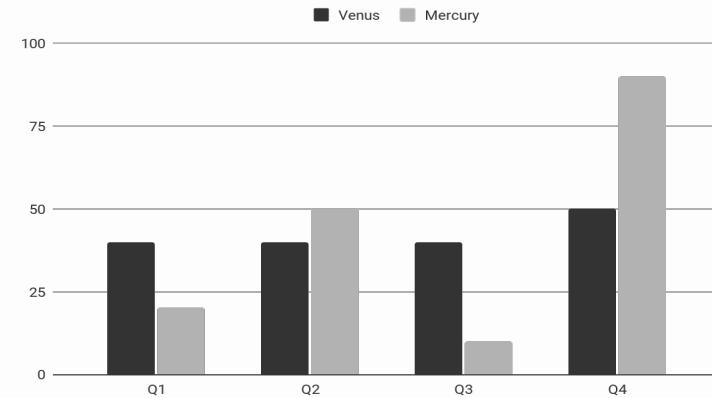
Phase 02

- Venus has a beautiful name and is the second planet from the Sun. It's terribly hot—even hotter than Mercury
- Saturn is a gas giant and has several rings. This planet is composed mostly of hydrogen and helium
- Earth is the third planet from the Sun and the only one that harbors life in the Solar System

Analysis & development

Mercury is the closest planet to the Sun and the smallest one in the Solar System. The planet's name has nothing to do with the liquid metal.

- The Sun is the star at the center of the Solar System
- Jupiter is the biggest planet in the entire Solar System
- Saturn is composed mostly of hydrogen and helium



Venus

Venus has a beautiful name

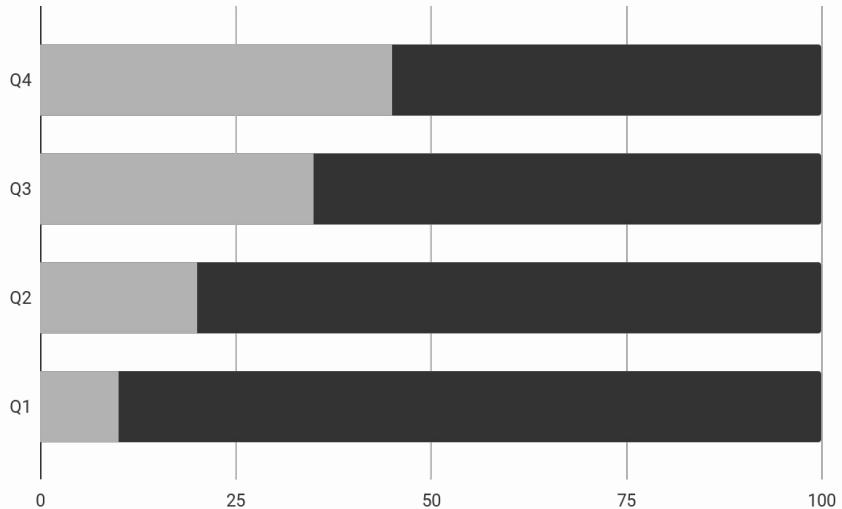


Mercury

Mercury is quite a small planet

Follow the link in the graph to modify its data and then paste the new one here. [For more info, click here](#)

Analysis of the results



Mercury

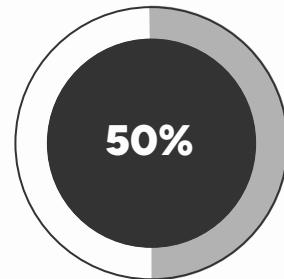
Mercury is quite a small planet

Venus

Venus has a beautiful name

Mars

Despite being red, Mars is a planet very cold



Follow the link in the graph to modify its data and then paste the new one here. [For more info, click here](#)

This is a map



Venus

Venus is the second planet from the Sun



Mercury

Mercury is the closest planet to the Sun



Mars

Despite being red, Mars is a cold place

Discussion

Mars & Earth

Despite being red, Mars is actually a cold place. It's full of iron oxide dust, which gives the planet its reddish cast. Earth is the third planet from the Sun and the only one that harbors life in the Solar System. This is where we all live:

- Ceres is located in the main asteroid belt
- The Moon is Earth's natural satellite
- Neptune is very far away from us
- Pluto is now considered a dwarf planet



Discussion 1

Mercury is the closest planet to the Sun and the smallest one in the Solar System. It's bit larger than the Moon



Discussion 2

Venus has a beautiful name and is the second planet from the Sun. Venus is a lot hotter than Mercury

Conclusions



Mars

Despite being red, Mars is actually a cold place. It's full of iron oxide dust



Mercury

Mercury is the closest planet to the Sun and the smallest one in the Solar System



Venus

Venus has a very beautiful name and is the second planet from the Sun



Neptune

Neptune is a big planet. It is the fourth-largest planet by diameter in the Solar System

Bibliographical references

Surname, A. (YEAR). Name of the source. Publisher

Surname, A. (YEAR). *Name of the source*. Publisher

Surname, A. (YEAR). Name of the source. Publisher

Surname, A (YEAR) Name of the source Publisher

Surname A (YEAR) Name of the source Publisher

Mockups

You can replace the images on the screen with your own work. Just right-click on them and select "Replace image"



Thanks!

Do you have any questions?

youremail@freepik.com

+34 654 321 432

yourwebsite.com



CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#)

Please keep this slide for attribution

Icon pack



Alternative resources

Here's an assortment of alternative resources whose style fits that of this template:

- Flat design geometric background



Resources

Did you like the resources in this template? Get them at these websites:

Icons

- [Icon Pack: Books | Outline](#)

Photos

- [Medium shot people at traveling agency](#)
- [Close up hand signing document](#)

Instructions for use

If you have a free account, in order to use this template, you must credit **Slidesgo** by keeping the **Thanks** slide. Please refer to the next slide to read the instructions for premium users.

As a Free user, you are allowed to:

- Modify this template.
- Use it for both personal and commercial projects.

You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs or visit our blog:
<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

Instructions for use (premium users)

As a Premium user, you can use this template without attributing Slidesgo or keeping the "Thanks" slide.

You are allowed to:

- Modify this template.
- Use it for both personal and commercial purposes.
- Hide or delete the "Thanks" slide and the mention to Slidesgo in the credits.
- Share this template in an editable format with people who are not part of your team.

You are not allowed to:

- Sublicense, sell or rent this Slidesgo Template (or a modified version of this Slidesgo Template).
- Distribute this Slidesgo Template (or a modified version of this Slidesgo Template) or include it in a database or in any other product or service that offers downloadable images, icons or presentations that may be subject to distribution or resale.
- Use any of the elements that are part of this Slidesgo Template in an isolated and separated way from this Template.
- Register any of the elements that are part of this template as a trademark or logo, or register it as a work in an intellectual property registry or similar.

For more information about editing slides, please read our FAQs or visit our blog:

<https://slidesgo.com/faqs> and <https://slidesgo.com/slidesgo-school>

Fonts & colors used

This presentation has been made using the following fonts:

Figtree

(<https://fonts.google.com/specimen/Figtree>)

Hanken Grotesk

(<https://fonts.google.com/specimen/Hanken+Grotesk>)

#333333

#fdfdfd

#e8e7e7

#b2b2b2

#777777

Storyset

Create your Story with our illustrated concepts. Choose the style you like the most, edit its colors, pick the background and layers you want to show and bring them to life with the animator panel! It will boost your presentation. Check out [how it works](#).



Pana



Amico



Bro



Rafiki



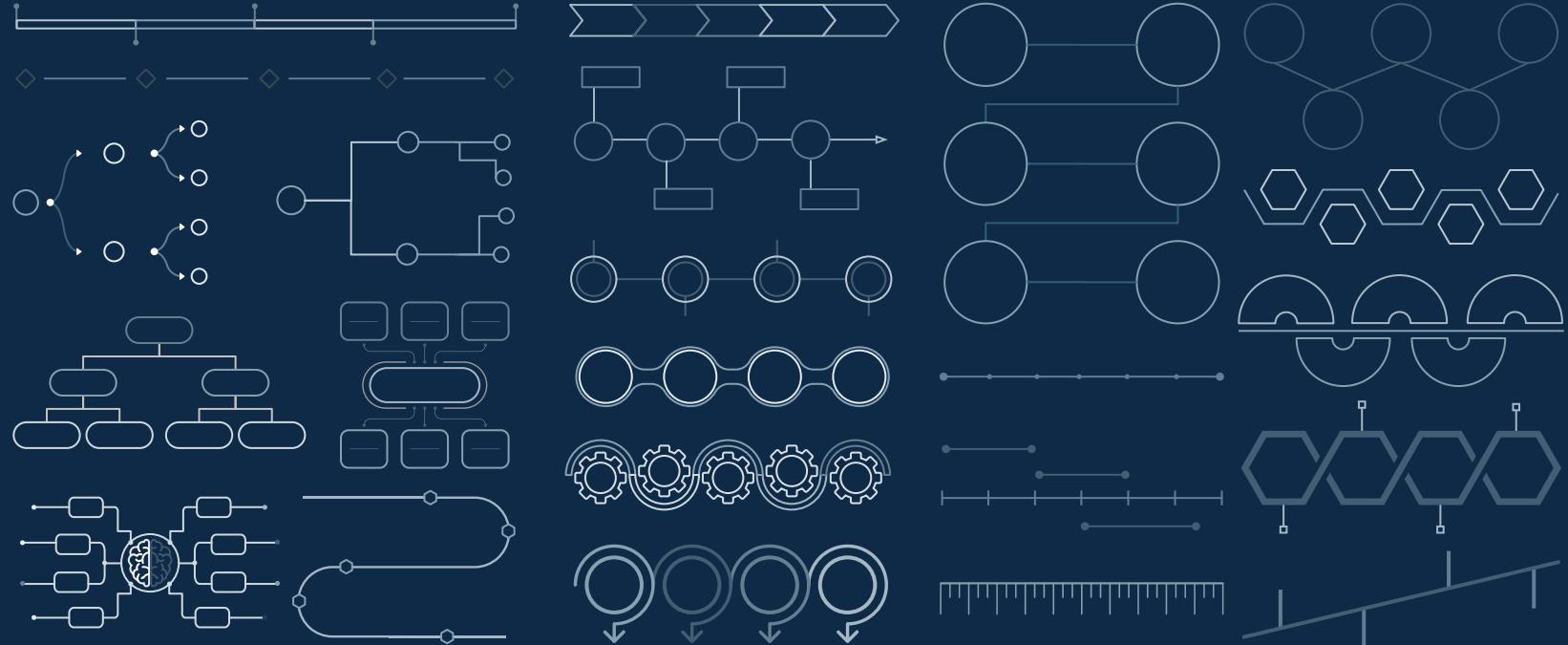
Cuate

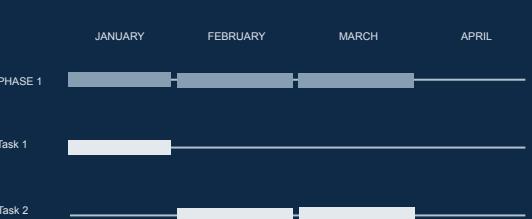
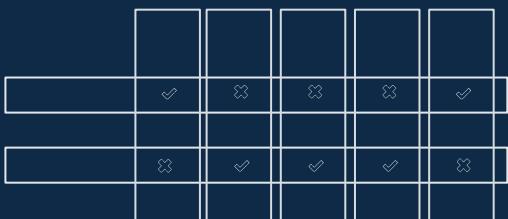
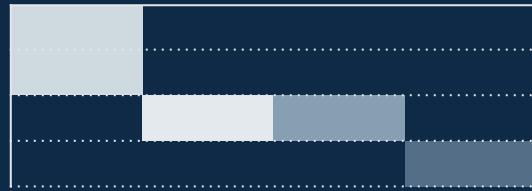
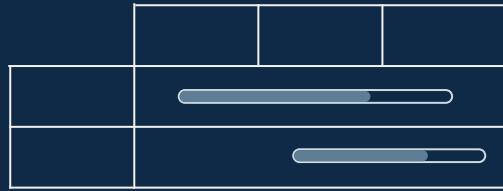
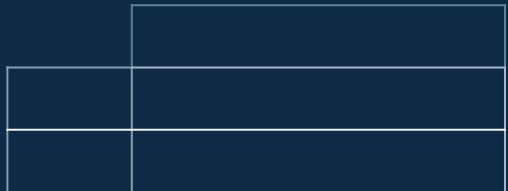
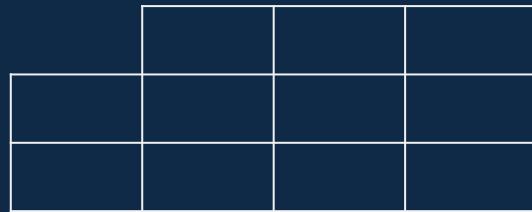
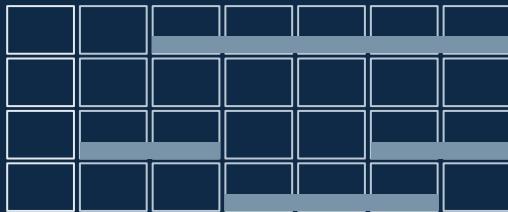
Use our editable graphic resources...

You can easily **resize** these resources without losing quality. To **change the color**, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done. You can also look for more **infographics** on Slidesgo.

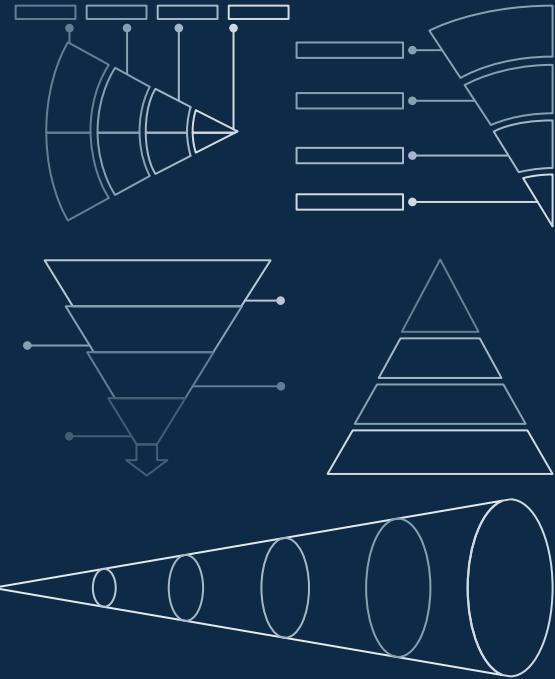
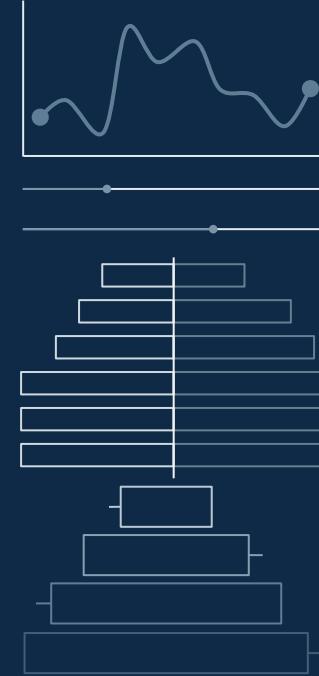
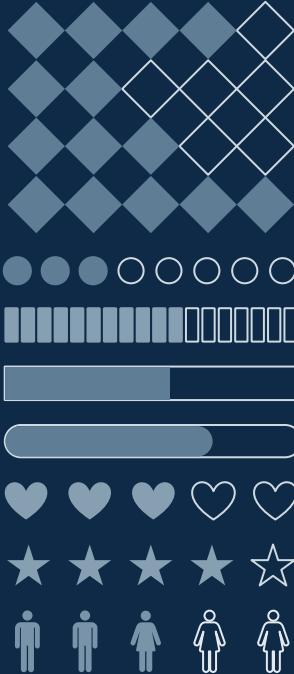
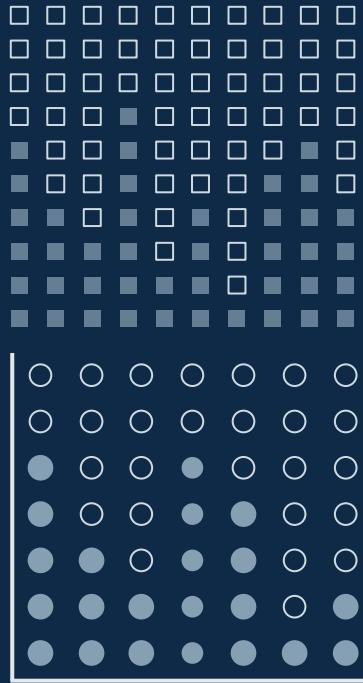












...and our sets of editable icons

You can **resize** these icons without losing quality.

You can **change the stroke and fill color**; just select the icon and click on the **paint bucket/pen**.

In Google Slides, you can also use **Flaticon's extension**, allowing you to customize and add even more icons.



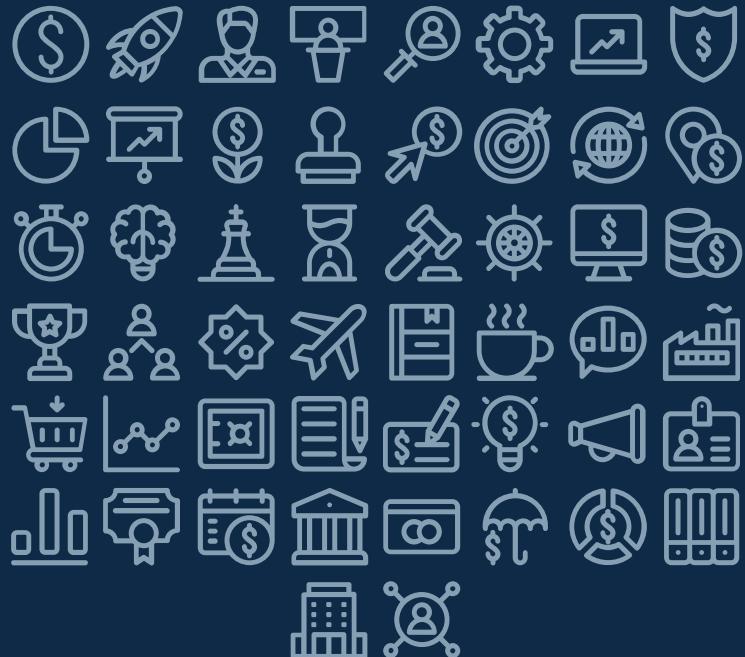
Educational Icons



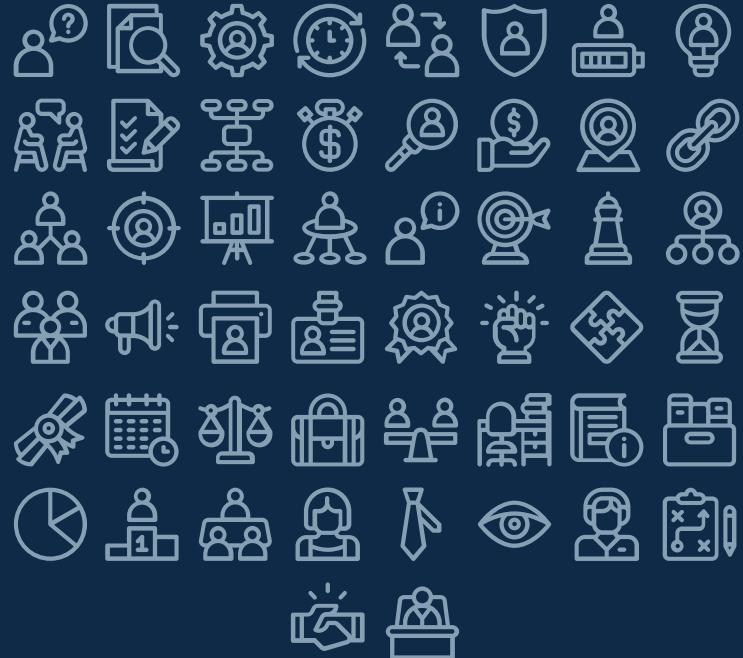
Medical Icons



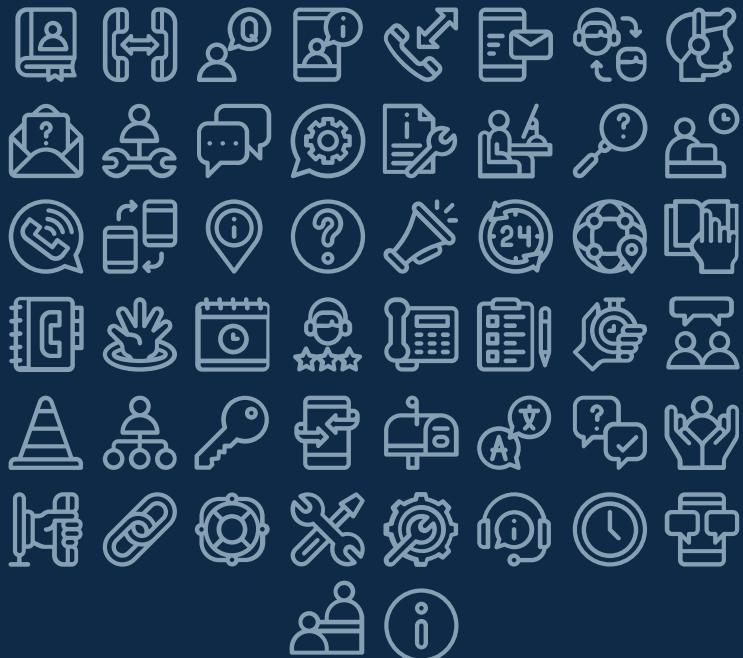
Business Icons



Teamwork Icons



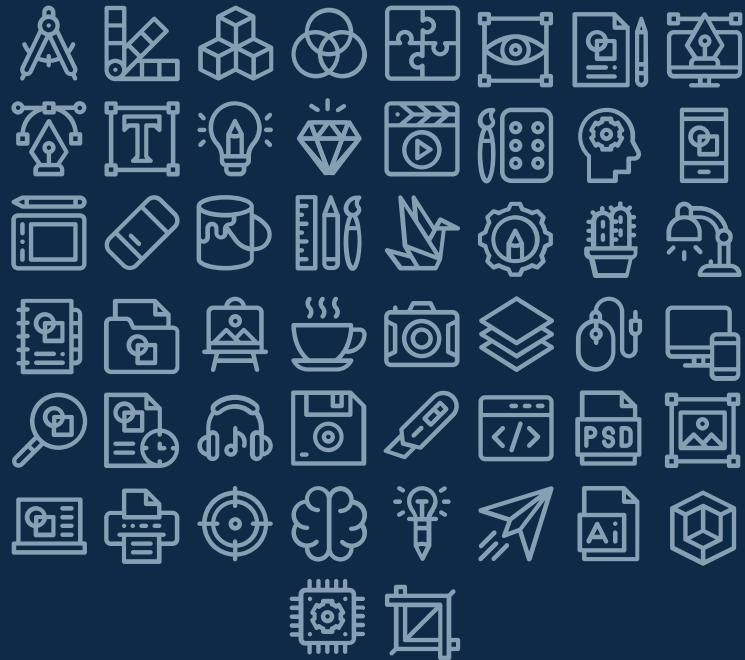
Help & Support Icons



Avatar Icons



Creative Process Icons



Performing Arts Icons



Nature Icons



SEO & Marketing Icons



