Distance and Related Notions in Graphs

Ekene Ezeunala

June 19th, 2022

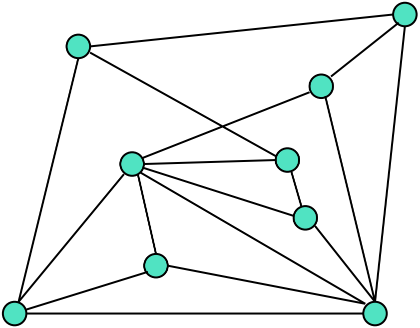
Outline

1. Introduction
2. Definitions
3. Representing Complex Graphs
4. Application to Music
5. References

Background from last talk

Definition (Connected graphs)

An undirected graph $\mathcal{G}$ for which there exists a path between every pair of vertices is said to be *connected*. Each maximal connected piece of a graph is called a *connected component*.

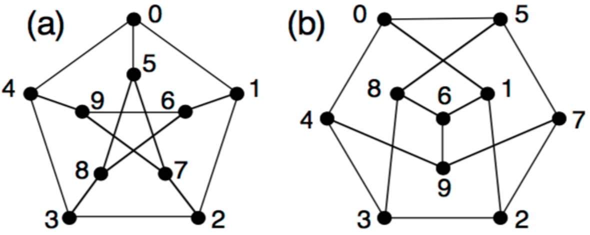


Graph isomorphisms

Definition (Isomorphism)

An *isomorphism of graphs* $\mathcal{G}_1$ and $\mathcal{G}_2$ is an edge-preserving bijection between the vertex sets of $\mathcal{G}_1$ and $\mathcal{G}_2$ such that any two vertices u and v of $\mathcal{G}_1$ are adjacent in $\mathcal{G}_1$ iff $f(u)$ and $f(v)$ are adjacent in $\mathcal{G}_2$.

Two graphs are said to be *isomorphic* if this bijection exists—loosely speaking, if you can draw $\mathcal{G}_1$ to look like $\mathcal{G}_2$.



Metrics

Definition (Metrics)

A *metric* $\mathbb{M}$ on a set X is a (distance) function $\mathbb{M} : X \times X \longrightarrow \mathbb{R}$ such that for all $x, y, z \in X$, the following three axioms hold:

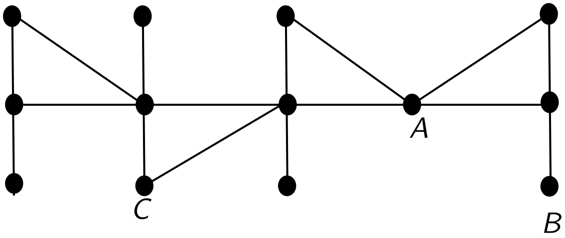
- $\mathbb{M}(x, y) = 0 \iff x = y$ (identity of indiscernibles)
- $\mathbb{M}(x, y) = \mathbb{M}(y, x)$ (symmetry)
- $\mathbb{M}(x, z) \leq \mathbb{M}(x, y) + \mathbb{M}(y, z)$ (triangle inequality)

- **Examples**: Euclidean: $\sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$, Manhattan: $\sum_{i=1}^{n} |x_i - y_i|$, Minkowski: $\sqrt[p]{\sum_{i=1}^{n} |x_i - y_i|^p}$ (with $p \geq 1$), etc.

Distance

Definition (Distance)

For a connected graph $\mathcal{G}$, the *distance* $d(u, v)$ from vertex u to vertex v is the length—ie number of edges—of a shortest u-v path in $\mathcal{G}$ (regardless of the number of alternative paths).



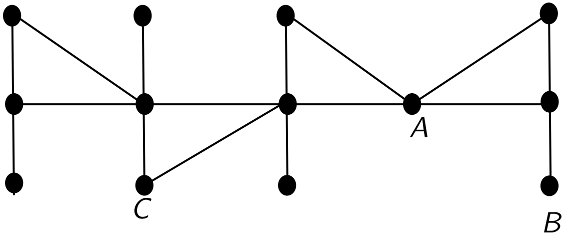
$$d(A, C) = d(A, B) = 2$$

Eccentricity

Definition (Eccentricity)

The greatest distance from a given vertex v to any other vertex x on a given graph is called the graph's *eccentricity*:

$$\mathrm{ecc}(v) = \max_{x \in V(\mathcal{G})} \{d(v, x)\}.$$



$\mathrm{ecc}(A) = \mathrm{ecc}(C) = 3 \qquad \mathrm{ecc}(B) = 5$

Radius, diameter, periphery, centre

For a given graph $\mathcal{G}$,

Definition (Radius, Diameter)

The radius $\mathrm{rad}(\mathcal{G})$ is the value of smallest eccentricity. The diameter $\mathrm{diam}(\mathcal{G})$ is defined as the value of greatest eccentricity.

Definition (Periphery, Centre)

The periphery is the set of vertices V such that $\mathrm{ecc}(V) = \mathrm{diam}(\mathcal{G})$. The centre is the set of vertices V such that $\mathrm{ecc}(V) = \mathrm{rad}(\mathcal{G})$.

Question: For what kind of graphs, if any, are $\mathrm{diam}(\mathcal{G})$ and $\mathrm{rad}(\mathcal{G})$ equal?

Two Properties

Theorem

Every connected graph $\mathcal{G}$ has $\mathrm{rad}(\mathcal{G}) \leq \mathrm{diam}(\mathcal{G}) \leq 2\mathrm{rad}(\mathcal{G})$.

Proof Outline.

We only need show the second inequality, $\mathrm{diam}(\mathcal{G}) \leq 2\mathrm{rad}(\mathcal{G})$. Choose vertices u and v such that $d(u, v) = \mathrm{diam}(\mathcal{G})$, and choose a vertex c in the centre. Then use the triangle inequality. $\square$

Theorem

Every graph of $n \geq 1$ vertices is isomorphic to the centre of some graph.

Motivation

Until now, we have employed mostly hand-wavy methods to understanding graphs and answering natural questions like

- Are two vertices connected by a sequence of edges (if not a single edge)?
- What is the minimum number of edges we need to traverse from one vertex to another?
- What is the minimum number of edges we need to traverse from any vertex on the graph to any other?
- How can we tell if a graph is connected?
- ...

These questions are not hard to answer for smaller graphs, but become more difficult as the number of edges and vertices grows (even with algorithms like Bellman-Ford).

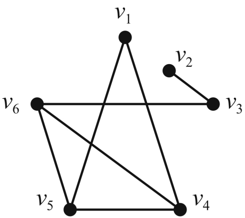
Adjacency matrix

Definition (Adjacency matrix)

Let $\mathcal{G}$ be a graph with vertices $v_1, v_2, \ldots, v_n$. Then the *adjacency matrix* of $\mathcal{G}$ is the $n \times n$ matrix A whose (i, j) entry, denoted by $[A]_{i,j}$, is defined by

$$[A]_{i,j} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are adjacent,} \\ 0 & \text{otherwise.} \end{cases}$$

Example:



$$A = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

A Few Properties

For adjacency matrices A, A_1, A_2 :

- For undirected graphs, A is symmetric about the major diagonal.
- On simple graphs, the main diagonal of A has all 0's.
- Given two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ with respective A_1 and A_2, if there is a permutation of the rows and columns of A_1 that gives A_2, then $\mathcal{G}_1$ and $\mathcal{G}_2$ are isomorphic.
- The sum of the vertices in a row or column equals the degree of the vertex it represents.

Walks and Adjacency Matrices

Theorem

For $\mathcal{G}$ a graph with vertices labelled as $v_1, v_2, \ldots, v_n$, A its corresponding adjacency matrix, and k a positive integer, the (i,j) entry of A^k is equal to the number of walks from v_i to v_j that use exactly k edges.

Proof Outline.

By induction. When $k = 1$, $[A_{i,j}] = 1$—only one edge is available. Let a_{ij} be the ij^{th} entry of A.

Suppose that the ij^{th} entry of A^k, b_{ij}, is the number of k-edge walks from v_i to v_j.The ij^{th} entry of A^{k+1} is then $\sum_{m=1}^{n} a_{im} b_{mj}$. Pick $a_{i1} b_{1j}$; this is the number of k-edge walk from v_1 to v_j times the number of 1-edge walks from v_i to v_1. This is in turn the number of $(k+1)$-edge walks from v_i to v_j. Now choose any other m: the argument still holds. $\square$

Matrix Sum

Definition (Matrix Sum)

Given a graph $\mathcal{G}$ of order n with adjacency matrix A, and given a positive integer k, define the matrix sum S_k to be

$$S_k = \mathbb{I} + A + A^2 + \ldots + A^k,$$

where $\mathbb{I}$ is the $n \times n$ identity matrix.

Motivation.

See previous theorem. □

Eccentricity, Radius, and Diameter

Theorem

Let $\mathcal{G}$ be a connected graph with vertices labelled $v_1, v_2, \ldots, v_n$, and let A be its corresponding adjacency matrix.

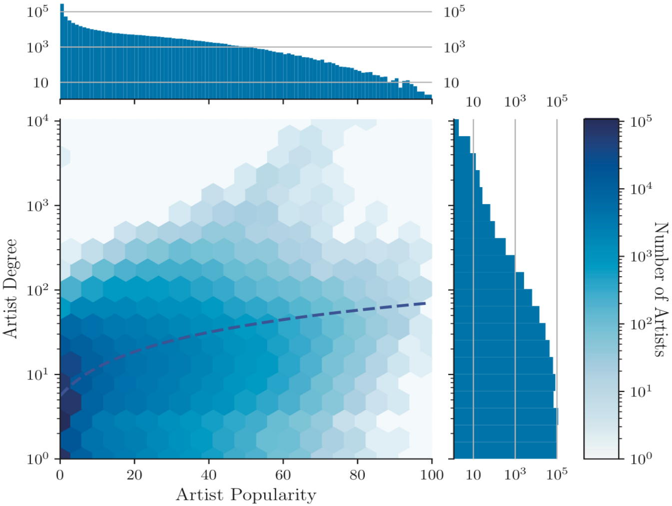
- *If k is the smallest positive integer such that row j of S_k contains no zeros, then $\mathrm{ecc}(v_j) = k$.*
- *If r is the smallest positive integer such that all entries of at least one row of S_r are positive, then $\mathrm{rad}(\mathcal{G}) = r$.*
- *If m is the smallest positive integer such that all entries of S_m are positive, then $\mathrm{diam}(\mathcal{G}) = m$.*

A network analysis of Spotify (by T. South et al)

- A network of all the artists on Spotify connected by who they worked with
- 1,250,065 artists (vertices on undirected graph)
- 3,766,631 collaborations (edges on undirected graph)
- Snowball sampling starting with Kanye West
- Get metadata on these artists (eg popularity, etc)
- Represent using adjacency matrix:

$$A = \begin{array}{c c} & \begin{array}{ccc} \text{Kanye} & \text{Drake} & \text{Taylor} \end{array} \\ \begin{array}{c} \text{Kanye} \\ \text{Drake} \\ \text{Taylor} \end{array} & \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{array}$$

Relative popularity



Eigenvector centrality

- Calculate the *eigenvector centrality*: This is just taking the eigenvector with the largest corresponding eigenvalue

$$Av = \lambda v$$

A

Filter popularity + eigenvector centrality

- Take the most popular contemporary artists

B

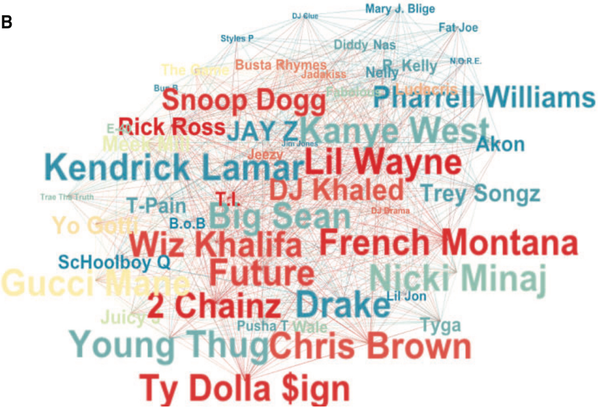
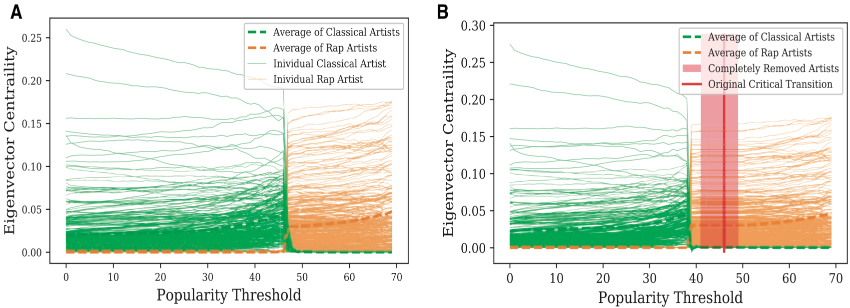


Figure: Dominance shifts from classical music to rappers.

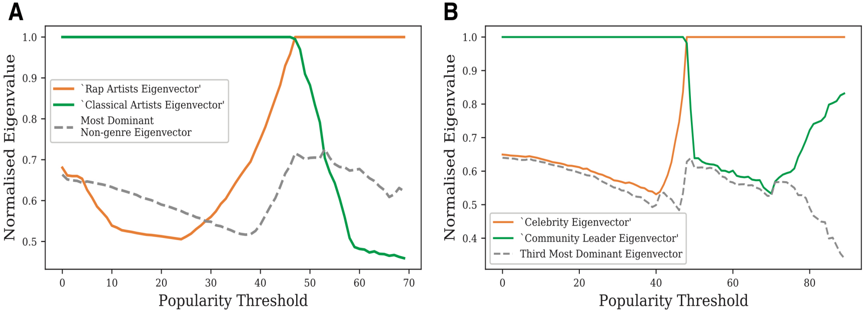
When does it all change?

- Take the network and chop off nodes with a popularity of 10 or less
- Nothing changes in that critical region, but the location of the transition is shifted
- Which means that the entire structure of the graph changes suddenly as we account for popularity!



Why? It's all in the eigenvalues

- The vectors representing classical dominance and rap dominance exist the entire time, but they change ranking:

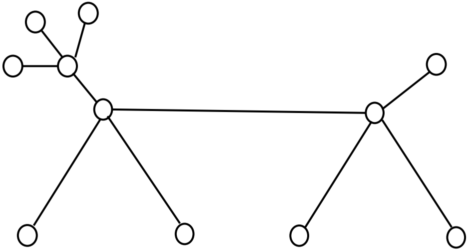


References and further reading

1. BLeversha, Gerry. "Combinatorics and graph theory, by John M. Harris, Jeffry L. Hirst, Michael J. Mossinghoff. Pp. 225.£ 24 (hb). 2000. ISBN 0 387 98736 3 (Springer-Verlag)." *The Mathematical Gazette* 86.505 (2002): 177-178.

2. South, Tobin, Matthew Roughan, and Lewis Mitchell. "Popularity and centrality in Spotify networks: critical transitions in eigenvector centrality." *Journal of Complex Networks* 8.6 (2020): cnaa050.

3. Bryan, Kurt, and Tanya Leise. "The $25,000,000,000 eigenvector: The linear algebra behind Google." *SIAM review* 48.3 (2006): 569-581.



This graph is well known for its bark.