Majority and Friendship Paradoxes

Majority Paradox

Example: Small town is considering a bond initiative in an upcoming election. Some residents are in favor, some are against.

Consider a poll asking the following two questions:

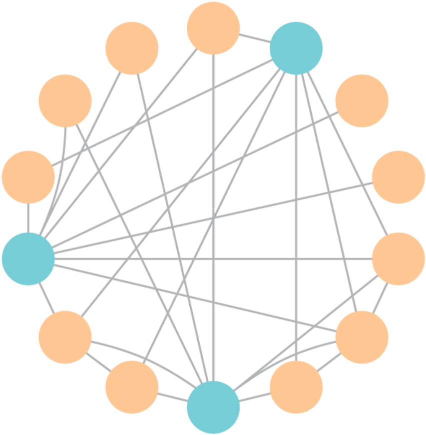
- Are you for or against the initiative?
- Do you think the initiative will pass?

Note that

- Answer to the first question depends on the opinion of the resident only.
- Answer to the second question depends on the opinion of the residents friends (and social/mainstream media they follow)

Upshot: Answers to the second question may be misleading/inaccurate

Social Network (from Wash. Post): Blue = For, Orange = Against



In this case, most residents are opposed to the initiative, but the majority of residents think the initiative will pass.

Friendship Paradox

Informal: On average, individuals in a social network have fewer friends than their friends do.

Formal: Let $G = (V, E)$ be an undirected graph. Define

- $A_1(G) =$ average degree of vertices in G
- $A_2(G) =$ average degree of the neighbors of vertices in G

Then $A_1(G) \leq A_2(G)$.

If G is a social network, then

- $A_1(G) =$ average number of friends of individuals in network
- $A_2(G) =$ average number of friends of friends in network

Proof of Friendship Paradox

Recall: $N(u)$ = neighbors of u in G, and $d(u) = |N(u)|$ = degree of u in G.

Let $|V| = n$. Then we can write

$$A_1(G) = \frac{1}{n}\sum_{u\in V} d(u) \quad \text{and} \quad A_2(G) = \frac{\sum_{u\in V}\left(\sum_{v\in N(u)} d(v)\right)}{\sum_{u\in V}|N(u)|}$$

Fact: The average $A_2(G) = \sum_{u\in V} d^2(u) \,/\, \sum_{u\in V} d(u)$

Rearranging, we find that

$$A_1(G) \le A_2(G) \;\longleftrightarrow\; \left(\frac{1}{n}\sum_{u\in V} d(u)\right)^2 \le \frac{1}{n}\sum_{u\in V} d^2(u)$$

Basic Lemma

To complete the proof, we appeal to a simple inequality.

Lemma: If $a_1, \ldots, a_n$ are real numbers then

$$\left(\frac{1}{n}\sum_{i=1}^{n} a_i\right)^2 \leq \frac{1}{n}\sum_{i=1}^{n} a_i^2$$

Proof: Homework.

In statistical terms this is a special case of the inequality $(EX)^2 \leq EX^2$ for a random variable X

Euler Paths

Euler Paths and Circuits

Definition: Let $G = (V, E)$ be a simple graph.

- An *Euler* path in G is a simple path that contains every edge in E.
- An *Euler* circuit in G is a simple circuit that contains every edge in E.

Example: Let G represent the map of a small town

- vertices $=$ intersections
- edges $=$ streets

Can postal worker deliver mail to residents of the town without walking any street twice?

Euler Paths and Circuits, cont.

Goal: Necessary and sufficient conditions for

- Euler paths in G
- Euler circuits in G

Punch line: There are simple conditions involving only the degree of the vertices in G

Euler Circuits and Even Degree

Theorem: Let $G = (V, E)$ be a connected multigraph with $|V| \geq 2$. Then G has an Euler circuit iff every vertex has even degree.

Proof sketch ($\Leftarrow$): Let $k = 0$ and let all edges in E be unmarked. Fix a vertex $u_0 \in V$ and proceed as follows.

(A) If all edges $e \in E$ containing u_k are marked, then *stop*. Otherwise

- Select and mark an edge $e_k = \{u_k, v\} \in E$.
- Let $u_{k+1} = v$ be the other endpoint of e_k.
- Increase k by one and return to (A)

Let $e_r = \{u_r, u_{r+1}\}$ be the last edge selected. Define the path

$$p = u_0, u_1, \ldots, u_{r+1}$$

Proof, cont.

Claim: Path p is a circuit from u_0 to u_0, that is, $u_{r+1} = u_0$.

Let $v \neq u_0$ be a vertex in p. Note:

- If $\#$ unmarked edges at v is ≥ 2, then when p enters v on one edge it can leave on another, reducing the number of unmarked edges by 2.

As deg(v) is even, the number of unmarked edges at v is always

- ≥ 2, in which case p will return to v and leave v
- $= 0$, in which case p will not return to v

Upshot: Path p cannot end at v, so it must terminate where it began, at u_0.

Proof, cont.

Next Steps: If the path p contains every edge in E, then it is the desired Euler circuit. Otherwise,

- Find a circuit $\tilde{p}$ in the subgraph G_1 of G generated by the edges $E_1 = E \setminus \{\text{edges in } p\}$.
- Splice p and $\tilde{p}$ together into a single circuit.
- Continue until all edges in E are used.

Euler Paths

Theorem: A connected graph $G = (V, E)$ has an Euler path, but not an Euler circuit, iff G has exactly two vertices of odd degree.

Hamilton Paths and Circuits

Hamilton Paths and Circuits

Definition: Let $G = (V, E)$ be a simple graph.

- A *Hamilton* path in G is a simple path that passes through every vertex exactly once.
- A *Hamilton* circuit in G is a Hamilton path that begins and ends at the same vertex.

Example: Let G represent the airline network for a region of the U.S.

A Hamilton path in G represents the flight itinerary of a salesperson who wishes to each city in the network once on a business trip.

Example: If $n \geq 3$ the complete graph K_n and the n-cycle C_n have Hamilton circuits.

Conditions for Hamilton Circuits

Theorem: Let $G = (V, E)$ be simple with $n \geq 3$ vertices. Then G has a Hamilton circuit if for every $u, v \in V$

$$\{u, v\} \notin E \;\Rightarrow\; \text{deg}(u) + \text{deg}(v) \geq n$$

Note: Theorem gives sufficient conditions for a Hamilton circuit. Simple necessary conditions are not known.

Planar Graphs

Planar Graphs

Definition: A graph $G = (V, E)$ is *planar* if it can be drawn in the plane in such a way that no two edges intersect (except at vertices). Any such drawing is called a *planar representation* of G.

Note

- A planar graph G can have many representations.
- Every planar representation divides the plane into a finite number of disjoint regions, one of which is unbounded.

Fact: If $e = \{u, v\}$ is *not* a cut edge of a graph G then there is a circuit c from u to u beginning with e.

Euler's Formula

Theorem: Let $G = (V, E)$ be a simple connected planar graph with $|E| = m$ and $|V| = n$. Then the number of regions r in *any* planar representation of G is $r = m - n + 2$.

Proof of Theorem: By strong induction on $m =$ number of edges in G.

Basis: If $m = 0$ then G is a single isolated vertex. In particular, G is connected, $n = 1$, and $r = 1 = m - n + 2$.

Induction: Let $m \geq 1$.

- Assume formula holds for graphs with at most $m - 1$ edges.
- Consider graph G with $|E| = m$ and $|V| = n$.
- Select and remove one edge $e = \{u, v\}$ from G

Proof of Euler's Formula

Case 1: Edge e is a cut edge of G.

Idea: $G \setminus e$ has connected components G_1 and G_2.

- G_1, G_2 are simple and connected.
- $m = m_1 + m_2 + 1$ and $n = n_1 + n_2$
- G planar implies G_1, G_2 planar
- The unbounded regions of G_1, G_2 overlap, but the other regions don't. Thus $r = r_1 + r_2 - 1$.

Proof of Euler's Formula

Case 2: Edge e is a not cut edge of G.

Idea: Using Fact, let $c = u, v, \ldots, u$ be the shortest simple cycle in G from u to u beginning with edge e.

- As c is simple and no two edges in G intersect, c encloses one or more regions of G.
- As c is a shortest cycle, it encloses a single region R.
- Removing e merges R with the region on the opposite side of e, reducing r by 1.

Examples

Example: A planar graph with $r = 1$ has no cycles.

Example: A connected planar graph G has 10 vertices and 15 edges. How many regions are there in a planar representation of G?

Example: The planar representation of a graph G has 4 regions. If G has 11 vertices, how many edges does it have?

Degree of a Region

Definition: Let R be a region of a connected simple planar graph. The *degree* of R is the number of edges traversed while walking along the boundary of R, with the interior to your left.

Note

- In the definition, edges can be traversed more than once
- The sum of deg(R) over the regions R of G counts every edge twice, so
$$\sum_{j=1}^{r} \text{deg}(R_j) \ = \ 2m$$

Degree of a Region, cont.

Fact: If $G = (V, E)$ is connected, planar, and simple with $|V| = n \geq 3$ then $|E| = m \leq 3n - 6$.

Corollary: The graph K_5 is not planar.

Corollary: If G is connected, planar, and simple, then there is a vertex $v \in V$ with $\deg(v) \leq 5$.

Graph Coloring

Graph Coloring

Definition: A *coloring* of a simple graph G is an assignment of colors to each vertex with the property that no adjacent vertices have the same color.

Definition: The *chromatic number* of a graph G, denoted by $\chi(G)$, is the minimum number of colors needed for a coloring of the graph.

Example: Graph G with wedding guests as vertices, and edges between guests who do *not* get along.

- Then $\chi(G) =$ minimal number of tables at the reception needed to guarantee that everyone seated at a table gets along.

Graph Coloring, cont.

Example: Graph G with classes as vertices, and edges between two classes if they have one or more students in common.

- Then $\chi(G) =$ minimal number of final exam time slots needed to avoid scheduling conflicts.

Example: If G is bipartite then $\chi(G) = 2$, and conversely.

Example: If K_n is the complete graph on n vertices, then $\chi(K_n) = n$.

Example: If C_n is an n-cycle with n even then $\chi(C_n) = 2$.

Famous Result

The Four Color Theorem: If G is planar then $\chi(G) \leq 4$.

Looking Backwards and Forwards

STOR 215: Introduction to Mathematical Reasoning

I. Logic: language of mathematical reasoning and higher mathematics

- compound propositions, logical operations, quantifiers

II. Basic objects of mathematical analysis

- sets, summations, functions

III. First look at specific subject areas

- number theory, combinatorics, graphs/networks

Where are these ideas used?

- linear algebra
- advanced calculus
- discrete mathematics
- econometrics
- probability
- optimization
- analysis of algorithms
- machine learning
- stochastic modeling

The End