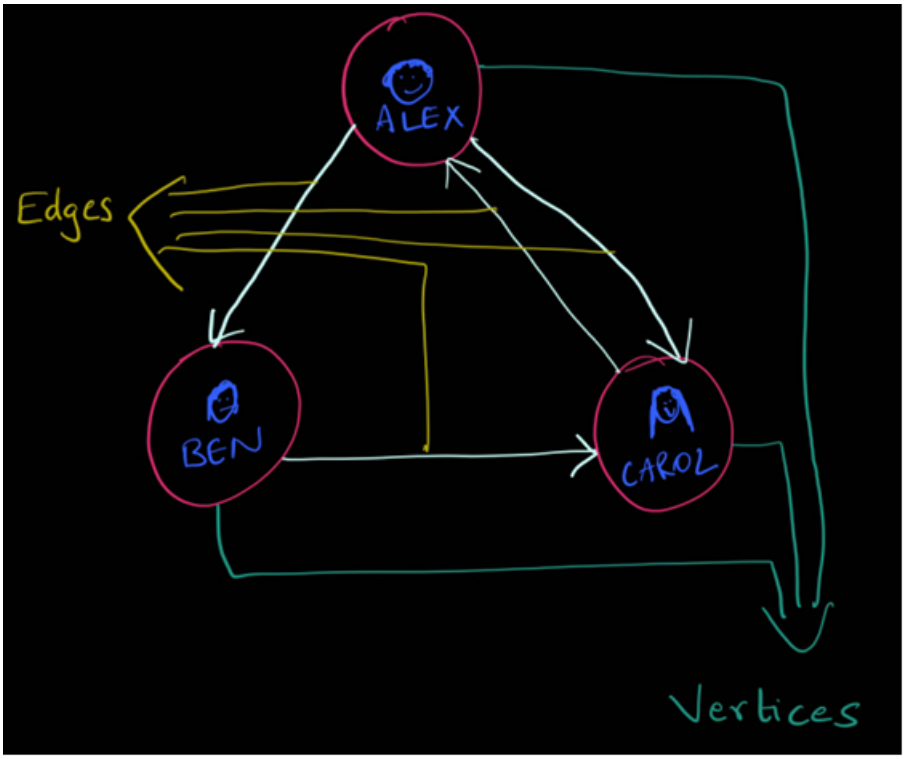
# Revision on Page Rank and Behavioral Intuition

Minor in AI, IIT Ropar

28th May, 2025

## 1 Session Overview and Philosophy

- **Feedback and Engagement:**
  - Monthly feedback encouraged for session improvement.
  - Emphasis on real-world intuition and application over mathematical formulas.
  - Revision covers both Module A and Module B, integrating concepts for comprehensive understanding.




## 2 Sweet Shop and Bookshop

### 2.1 Sweet Shop Scenario

- Shelves organized by sweet type (laddoos, gulab jamun, etc.).

- Customers sample sweets, starting with favorites or visually appealing options.
- Decision-making influenced by variety, arrangement, and personal preference.
- Shop offers free samples to encourage tasting and selection.
- Behavior patterns: customers may try many varieties, but typically return to favorites or most appealing options.

### 2.2 Bookshop Scenario

- Books arranged by genre, then by author's last name within each genre.
- Customers browse genres, read sample pages, and may switch genres if not satisfied.
- Behavior patterns: customers may stay in one genre for a while, then move to another if nothing appeals.

### 2.3 Common Observations

- **Arrangement:** Items grouped by similarity (sweets by type, books by genre).
- **Variety:** Multiple options available, leading to decision fatigue.
- **Sampling:** Customers try samples or read excerpts before making a choice.
- **Segmentation:** Popular or best-selling items are prominently displayed.
- **Transition:** Customers may move between sections based on interest or satisfaction.

## 3 Scientific Approach: Probability and Observation

- **Observation:**
  - Researchers observe behavior in shops, noting time spent and transitions between shelves.
  - Example: 80% chance a person stays in the same section, 20% chance they move to another.
- **Probability in Behavior:**
  - Used to model real-world scenarios (e.g., bookshop browsing, web page navigation).
  - Introduces the concept of the **damping factor** (e.g., 85% chance to stay, 15% to jump).

# 4 Page Rank Algorithm: Intuition and Application

## 4.1 Graph Theory Basics

- **Graph:** Set of nodes (vertices) and edges (connections).
- **Directed Graph:** Edges have direction (e.g., from node A to node B).
- **In-degree:** Number of edges coming into a node.
- **Out-degree:** Number of edges going out from a node.

## 4.2 Page Rank Intuition

- Inspired by real-world browsing behavior (e.g., sweet shop, bookshop).
- **Random Surfer Model:** User starts at a random page, follows links, or jumps to a new page.
- **Damping Factor:** Probability the user stays on the current page (e.g., 0.85) vs. jumps (0.15).

## 4.3 Algorithm Steps

1. **Initialization:** All pages start with equal rank.
2. **Iteration:**
   - Rank is updated based on in-links and out-links.
   - Formula:
   $$PR(A) = \frac{1-d}{N} + d\sum_{i} \frac{PR(T_i)}{C(T_i)}$$
     - $PR(A)$: Page Rank of page A
     - $d$: Damping factor
     - $N$: Total number of pages
     - $T_i$: Pages pointing to A
     - $C(T_i)$: Out-degree of $T_i$
3. **Convergence:** Repeat until ranks stabilize.

## 4.4 Implementation

- **NetworkX Library:** Used to implement graph structures and page rank in Python.
- **Trap State:** Node with no outgoing edges (user gets stuck).

# 5 Key Takeaways and Exam Tips

- **Intuition Over Math:**
  - Focus on understanding real-world scenarios and behavioral patterns.
  - Math details can be revisited in recorded sessions if needed.
- **Damping Factor:**
  - Represents the probability of staying vs. jumping in a graph.
  - Typical value: 0.85 (stay), 0.15 (jump).
- **Graph Representation:**
  - Nodes and directed edges model relationships.
  - In-degree and out-degree are crucial for ranking.
- **Implementation:**
  - Use libraries like NetworkX for quick implementation.
  - Graph data structures are widely applicable in AI and data science.
- **Exam Questions:**
  - May ask to explain the damping factor in the context of real-world behavior.
  - Could involve identifying in-degree and out-degree in a graph.
  - May require understanding how the algorithm converges and its real-world relevance.

# 6 Additional Insights

- **Live Sessions and Interaction:**
  - Attending live sessions is encouraged for better understanding and interaction.
  - Feedback from students shapes future sessions.
- **Six Degrees of Separation:**
  - Concept from social networks: everyone is connected to anyone else through a chain of six people.
  - Used to illustrate the power of graph theory in understanding connections.
- **Real-World Data Application:**
  - Example: Metro Interstate traffic data set for regression analysis.
  - Encouragement to experiment with code and data for deeper understanding.

# 7 Summary Table: Page Rank Concepts

| Concept | Description |
|---|---|
| Graph | Nodes and edges representing relationships |
| In-degree | Number of edges coming into a node |
| Out-degree | Number of edges going out from a node |
| Damping Factor | Probability user stays on current page (e.g., 0.85) |
| Page Rank | Algorithm to rank nodes based on link structure and probability |
| NetworkX | Python library for graph operations and page rank implementation |
| Trap State | Node with no outgoing edges |