# Social Hub for UIC

Pratyay Banerjee
Department of Computer Science
University of Illinois, Chicago
UIN: 657166077
pbane8@uic.edu

Niharika Belavadi Shekar
Department of Computer Science
University of Illinois, Chicago
UIN: 675234184
nbela@uic.edu

## I. Problem Description

The University of Illinois Chicago (UIC) is primarily a commuter school, where a majority of students travel to campus for classes and leave shortly after. This dynamic reduces the opportunities for students to build meaningful social connections, find study partners, or participate in campus activities outside of class hours. As a result, many students experience challenges in forming friendships and engaging with the broader campus community.

This project aims to address that gap by developing a **Social Hub** platform for UIC students. The proposed system will leverage machine learning to recommend peers with similar academic backgrounds, interests, or course enrollments, thereby facilitating both social and academic connections. The system will primarily consist of two features:

- **Friend Finder:** A recommendation system that identifies potential peers with shared attributes such as major, year, or personal interests.
- **Study Buddy Matcher:** A feature that suggests classmates who are enrolled in the same courses or share overlapping academic goals, enabling students to form study groups.

The recommendation engine will incorporate Graph Neural Networks (GCN/GraphSAGE) to model students and their connections as a graph, Collaborative Filtering to refine recommendations based on user feedback and preferences, and embedding-based representation learning to capture similarities in student bios and interests.

As an additional extension, the system may also provide event recommendations for on-campus or Chicago-based activities aligned with student interests. This feature will be pursued if time permits, as the primary focus remains on enabling peer-to-peer connections.

By combining socially relevant goals with advanced machine learning methods, this project seeks to demonstrate the application of modern recommender system techniques to a real-world problem faced by commuter students.

## II. Datasets

To develop and evaluate the Friend Finder module, we will leverage publicly available student profile datasets and, if needed, generate synthetic data tailored to the UIC context.

### A. Students' Social Network Profile Clustering Dataset (Kaggle)

This dataset contains profiles of 15,000 high school students collected from a social networking platform (2006–2009). It includes demographic attributes (graduation year, gender, age, number of friends) and interest-related features extracted through text mining (counts of 37 dominant words such as "football" or "shopping"). This dataset is suitable for clustering students by interests and studying demographic correlations.

### B. Synthetic Student Profiles Dataset (Kaggle)

This dataset provides structured student profiles with demographic details (age, sex, country, state), academic information (major, GPA, year), hobbies, unique qualities, and short narratives. It offers a more diverse representation of student attributes, which can be valuable for building recommendation systems that combine structured and textual data.

### C. Dummy Synthetic Data

In addition to these datasets, we may create synthetic data that reflects attributes more specific to UIC commuter students (e.g., commute time, club memberships, or course enrollments). This allows us to simulate realistic profiles and tailor the dataset to the goals of the Social Hub project.

## III. Process and Implementation Plan

The proposed system will be developed in two main modules: **Friend Finder** and **Study Buddy Finder**, complemented by an interactive **AI Agent**.

### A. Data Preparation

- Construct or simulate student profiles with demographic, academic, and interest-based attributes.
- Encode categorical variables (e.g., major, year, clubs) into embeddings, and generate semantic representations for textual bios using pretrained language models such as Sentence-BERT.
- For the Study Buddy module, represent course enrollments as a bipartite graph of students and courses.

### B. Baseline Methods

- **Friend Finder:** Apply clustering algorithms (e.g., k-means, spectral clustering) on student embeddings and recommend peers from the same or nearest clusters.
- **Study Buddy Finder:** Match students based on course overlap and compatible study schedules, supplemented by clustering to group students with similar patterns.

### C. Advanced Models

- **Graph Neural Networks (GCN, GraphSAGE):** Train on the student graph to learn node embeddings and perform link prediction, identifying potential friendships or study partnerships.
- **Collaborative Filtering:** Use implicit feedback (e.g., acceptance/rejection of partner suggestions) to refine compatibility predictions.
- **Complementary Matching (Study Buddy):** Integrate GPA-based pairing strategies (e.g., matching higher GPA students with lower GPA peers) to promote peer-to-peer learning.

### D. Recommendation Generation

- Generate ranked lists of top-k matches for each student.
- Ensure both relevance (course overlap, shared interests) and diversity (different majors or backgrounds).

### E. AI Agent for Natural Language Queries

- Implement an AI Agent to enable students to interact with the system through natural language.
- Example query: *"I am on campus. I have an exam for CS 401 in a week. Find me a study buddy."*
- The agent will parse the query, extract intent (course, location, time frame), and interact with the trained models (clustering, GNNs, collaborative filtering) to generate personalized recommendations.
- Infrastructure: The agent will be deployed on **MCP (Model Context Protocol)** servers or **ADK (Agent Development Kit)** frameworks, enabling seamless integration between natural language interfaces and the underlying ML models.

### F. Evaluation

- **Baseline vs. Advanced Models:** Compare clustering and GNN-based approaches using Precision@k, Recall@k, and diversity.
- **Scalability:** Evaluate GraphSAGE for inductive performance on new students.
- **Agent Performance:** Assess query parsing accuracy and overall user satisfaction.

## IV. Related Work and Distinction of Proposed Work

Friend and study partner recommendation has been widely studied in both social networks and educational systems. Early approaches relied on heuristics such as common neighbors and Jaccard similarity [6], while Parveen and Varma [1] compared classical ML algorithms (K-Means, Random Forest, etc.) for friend recommendation. Collaborative Filtering approaches [5] and embedding-based methods [2] improved personalization but suffer from cold-start issues.

Recently, GNN-based methods have advanced the field. Hamidi et al. [3] benchmarked GCN, GraphSAGE, and LightGCN, showing strong efficiency, while Li et al. [4] surveyed GNN-based recommenders, emphasizing their ability to capture higher-order effects. In education, clustering-based study group formation [7] and collaborative filtering for MOOCs have been explored, but often neglect broader social attributes.

The proposed work differs in three ways:

1) **Commuter School Context:** Unlike MOOCs or generic social networks, our system directly addresses the isolation challenges of commuter students at UIC.
2) **Hybrid Approach:** We integrate clustering, collaborative filtering, embeddings, and GNNs, rather than relying on a single paradigm.
3) **AI Agent:** We introduce an MCP/ADK-based natural language interface that allows students to request recommendations dynamically (e.g., "find me a study buddy for CS 401").

## V. Literature Survey

### A. Clustering-Based Approaches

Parveen and Varma [1] investigated clustering and traditional machine learning algorithms (K-Means, Decision Trees, Naive Bayes, Logistic Regression, Random Forest) for friend recommendation on social media. Their results showed Random Forest achieved the highest accuracy, while clustering methods like K-Means produced broad, non-personalized groups. These approaches are useful as baselines but limited in scalability and personalization.

### B. Collaborative Filtering and Embedding Approaches

Collaborative Filtering (CF) methods, such as matrix factorization and neighborhood models, have been widely studied for recommendation but face cold-start limitations. Embedding-based methods extend CF by mapping users and items into continuous vector spaces, capturing latent semantic similarities. Zhao et al. [2] reviewed embedding-based recommender systems and highlighted their strengths in sparse datasets, though scalability and interpretability remain challenges. Embedding-based CF is particularly relevant for academic peer matching, where both structural and semantic signals are important. Vema and Devannd [5] also explored CF in user–item systems, demonstrating its utility in educational and social contexts.

Recent advances have integrated deep learning with collaborative filtering. Zaremarjal and Yiltas-Kaplan [8] proposed a semantic collaborative filtering approach using CNNs, demonstrating how convolutional architectures can extract meaningful features from user-item interaction patterns. Li et al. [9] surveyed deep neural network approaches in collaborative filtering, covering autoencoders, neural matrix factorization, and attention mechanisms, showing that deep learning methods consistently outperform traditional CF in capturing complex user preferences. Raza et al. [10] provided a comprehensive review transitioning recommender systems from theory to practice, examining real-world deployment challenges including scalability, interpretability, and user trust. Roy and Dutta [11] conducted a systematic review of recommender systems across domains, identifying key challenges such as data sparsity, cold-start problems, and the need for context-aware recommendations, particularly relevant for educational settings.

### C. Graph Neural Networks (GNNs)

Recent work has applied GNNs to recommendation as link prediction tasks. Hamidi et al. [3] benchmarked GCN, GraphSAGE, NGCF, and LightGCN, finding LightGCN competitive in efficiency and accuracy. Li et al. [4] surveyed GNN-based recommendation in social networks, emphasizing their ability to capture higher-order neighborhood effects beyond CF. Kumar and Vineela [6] further applied graph mining techniques for friend recommendation on social media, showing that graph-based methods outperform traditional clustering in capturing hidden structural similarities.

## VI. Preliminary Plan (Milestones)

- **Project Planning and Proposal:** Defining project scope based on course requirements and the specific challenges faced by UIC commuter students. Identify key features (Friend Finder, Study Buddy Matcher) and create the project proposal document.
- **Dataset Collection and Preprocessing:** Acquiring publicly available student profile datasets from Kaggle (Students' Social Network Profile Clustering Dataset and Synthetic Student Profiles Dataset). Generating synthetic data specific to UIC context including commute times, club memberships, and course enrollments. Cleaning datasets by handling missing values and standardizing attributes.
- **Data Preparation and Feature Engineering:** Encoding categorical variables (major, year, clubs) into embeddings using appropriate techniques. Generate semantic representations for textual student bios using pretrained language models such as Sentence-BERT. Construct bipartite graph representation of students and courses for the Study Buddy module.
- **Baseline Model Implementation:** Implementing clustering algorithms (k-means, spectral clustering) for Friend Finder recommendations. Developing course overlap matching system for Study Buddy Finder. Evaluating baseline performance using Precision@k, Recall@k, and diversity metrics.
- **Advanced Model Development:** Implementing Graph Neural Networks (GCN/GraphSAGE) for student graph modeling and link prediction. Integrating Collaborative Filtering for recommendation refinement based on user feedback. Developing complementary matching strategies incorporating GPA-based pairing for Study Buddy feature.
- **AI Agent Integration:** Designing and implementing natural language query interface for student interaction. Deploying agent on MCP (Model Context Protocol) or ADK (Agent Development Kit) frameworks. Developing query parsing and intent extraction modules to interact with underlying ML models.
- **System Evaluation and Testing:** Comparing baseline versus advanced model performance across multiple metrics. Evaluating GraphSAGE scalability with new student profiles. Assessing AI agent query parsing accuracy and user satisfaction through testing scenarios.
- **Intermediate Project Report:** Documenting progress at mid-semester checkpoint, including dataset preparation results, baseline model performance, and initial challenges encountered. Incorporating feedback for remaining development phases.
- **Final Project Presentation:** Preparing presentation covering project motivation, methodology, system architecture, experimental results, and demonstrations of Friend Finder and Study Buddy features. Present findings to class with visual aids and performance comparisons.
- **Final Project Report:** Completing comprehensive final report documenting entire project lifecycle, including literature review, methodology, implementation details, experimental results, system limitations, and future work directions.

## References


[1] Parveen, R. & Varma, N.S. (2021). Friend's recommendation on social media using different algorithms of machine learning.

[2] Zhao, X., Wang, M., & Zhao, X. (2022). Embedding in Recommender Systems: A Survey.

[3] Hamidi, C., Gaou, M., Tribak, H., & Gaou, S. (2024). Recommender Systems with Graph Neural Networks: A Comparative Study.

[4] Li, X., Sun, L., Ling, M., & Peng, Y. (2023). A survey of graph neural network based recommendation in social networks.

[5] Vema, N., & Devannd, B.A. (2019). User–item recommendation system (UIRS) using collaborative filtering.

[6] Kumar, K.N., & Vineela, K. (2020). Friend Recommendation Using Graph Mining on Social Media.

[7] Wang, J., Yang, D., & Wen, M. (2017). Investigating how student's cognitive behaviors in MOOC discussion forums affect learning achievements. *International Educational Data Mining Conference*.

[8] Zaremarjal, A.Y. & Yiltas-Kaplan, D. (2021). Semantic collaborative filtering recommender system using CNNs. *2021 8th International Conference on Electrical and Electronics Engineering (ICEEE)*, pp. 254-258. IEEE.

[9] Li, P., Noah, S.A.M. & Sarim, H.M. (2024). A survey on deep neural networks in collaborative filtering recommendation systems. *arXiv preprint arXiv:2412.01378*.

[10] Raza, S., Rahman, M., Kamawal, S., Toroghi, A., Raval, A., Navah, F. & Kazemeini, A. (2024). A comprehensive review of recommender systems: Transitioning from theory to practice. *arXiv preprint arXiv:2407.13699*.

[11] Roy, D. & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, 9(1), p.59.