Network Analysis of Winnipeg's 2025 Primary Transit Network Overhaul

COMP 4710 – Introduction to Data Mining

(Winter 2026) | **Project Proposal – Group 11**

1. Group Roles

Member	Role	Primary Responsibilities
Ahmed Hasan	**Group Lead**, Data Engineering	Project coordination, GTFS preprocessing, GitHub management
Cathy Li	Network Analysis	Graph construction, centrality metrics, clustering, connectivity analysis
Stephenie Michael	Visualization & Reporting	Data visualization, report drafting, meeting minutes, interactive GIS maps
Sudipta Sarker	External Data & Coverage Analysis	Neighbourhood data integration, population joins, coverage gap analysis, outlier detection

Meeting Schedule: The group will hold weekly meetings, coordinated via Lettuce Meet (https://lettucemeet.com/) with availability polls distributed on Sunday. Additional meetings will be scheduled before major submission deadlines to conduct final reviews and obtain consent declarations

---

2. Project Topic and Rationale

Motivation: On June 29, 2025, Winnipeg Transit launched the Primary Transit Network (PTN), representing the most significant overhaul of the city's bus system in decades. Following earlier incremental changes such as the introduction of the Blue Rapid Transit Line, this redesign replaced the traditional route structure with a spine-and-feeder model intended to improve reliability, simplify service, and strengthen connectivity across the city. Our project focuses on analyzing the structure and service coverage of this new network. Using Winnipeg Transit's GTFS schedule feed as our core dataset, we will model the bus system as a directed weighted network (stops as nodes, route connections as edges) and compute both service-level metrics (route frequency, connectivity, pass-ups) and network-level metrics (centrality, clustering) to identify which hubs are most crucial to mobility and which areas may remain weakly served.

Relevance: This topic is relevant because public transit access and reliability directly affect commuting and day to day quality of life in Winnipeg. Not to mention all of the members of our group are dependent on the bus in our daily life. This topic is especially relevant in Winnipeg because winter conditions make reliable bus service crucial; no one wants to wait in the cold for a late or infrequent bus.

Course Alignment: The project aligns strongly with the objectives of the course by requiring an end-to-end data mining workflow. This includes data selection and preprocessing (GTFS parsing and external data integration), visualization (maps and network plots), analysis (graph and network analytics, clustering, and outlier detection), and evaluation (quantifying connectivity patterns and validating

insights across datasets). These components reflect key course learning outcomes in data preprocessing, pattern mining, clustering, graph and network mining, and visual analytics.

Overall Goal: While the scope of external variables and the final set of metrics may evolve as we explore data availability and quality throughout the course, our overall goal is to produce actionable, **data-driven** insights into how well the current Primary Transit Network connects different parts of Winnipeg and where the most meaningful opportunities for improvement may lie

---

3. Rough Plan and Approach

Our approach combines publicly available GTFS transit and census data with graph-based analysis techniques to evaluate the structure and coverage of Winnipeg's new Primary Transit Network. The following subsections outline our data sources, methodology, tools, and anticipated challenges.

Datasets: We will use five publicly available datasets. The core dataset is Winnipeg Transit's GTFS feed, which contains approximately 5,000 stops and 80+ routes. We will supplement this with operational metrics (pass-ups, on-time performance) and geographic context (neighbourhood boundaries, census population) to enable coverage analysis.

Dataset	Source	Description	Usage
Winnipeg Transit GTFS	Winnipeg Transit API + TransitLand	Static schedule feed containing stops, routes, trips, stop times and , calendar	Build network; compute travel times, frequency, and connectivity metrics
Transit Pass-Ups	Winnipeg Open Data	Daily records of pass-up events logged by drivers when buses can't board passengers	Crowding and unmet demand proxy
On-Time Performance	Winnipeg Open Data	Scheduled vs. actual arrival times by route and time period	Reliability analysis by route, stop, and time window
Passenger Counts	Winnipeg Open Data	Estimated daily ridership by route	Demand proxy; equity interpretation
Neighbourhood Areas (NCA)	Winnipeg Open Data	Neighbourhood boundary polygons	service metrics by area; identify coverage gaps
Community Areas (CCA)	Winnipeg Open Data	Community area polygons grouping multiple neighbourhoods	Alternative spatial aggregation unit
2021 Census	Statistics Canada	Population, demographic and density data	Population joins; equity-of-access analysis

Methodology: Our workflow proceeds through five stages, from raw data acquisition through interactive visualization:

A. **Data Acquisition & Preprocessing:** Download the GTFS feed and supplementary datasets. Parse feed to extract schedules. Clean and validate coordinates, remove duplicate records, and standardize time formats.
B. **Graph Construction:** Model the transit network as a directed weighted graph using NetworkX. Nodes represent stops (with latitude and longitude attributes), and edges represent consecutive stop connections weighted by scheduled travel time and service frequency.

C. **Network Analysis:** Compute centrality metrics (degree, betweenness, and closeness) to identify critical hubs. Apply Louvain community detection to discover service clusters. Analyze connectivity patterns and key transfer points.
D. **Coverage & Outlier Analysis:** Join neighbourhood polygons with stop locations to compute service density per area. Identify underserved neighbourhoods using outlier detection. Correlate service levels with population density from census data
E. **Visualization:** Create interactive GIS maps using Folium and GeoPandas showing network structure, centrality heatmaps, service clusters, and coverage gaps. Generate static network plots for inclusion in the final report.

Tools: Our technical stack centers on Python for reproducible data mining workflows, using Pandas and NumPy (efficient data manipulation and cleaning), NetworkX (graph construction and centrality computation), and GeoPandas with Shapely (spatial joins and geographic operations). Visualization will be done using Folium (interactive web-based maps) and Matplotlib (static figures for the report). If time permits, we may build a lightweight interactive dashboard using Astro with React (fast static site generation with dynamic components) to explore precomputed results; this will be treated as an optional MVP. Version control is managed through Git and GitHub (collaborative development and reproducibility), with team coordination via Lettuce Meet (scheduling) and Telegram (communication).

Anticipated Challenges: GTFS data may contain missing or inconsistent records; we will implement validation checks and document cleaning decisions. Spatial joins may have edge cases due to boundary overlaps; we will verify results visually on maps. Graph metrics may be computationally expensive; we will use sparse representations and batch processing. To prevent scope creep, we will lock core metrics early and treat additional features as optional.

---

4. Task Allocation And Timeline

The following timeline outlines our planned task allocation by submission deadline. Responsibilities may shift as we encounter data challenges or refine our scope, and any changes will be documented in progress reports.

Stage	Deadline	Ahmed	Cathy	Stephenie	Sudipta
Proposal	Jan 20	Scope, feasibility, submit	Network methods review	Draft Sections	Data research
Progress Report 1	Feb 5	GTFS preprocessing; frequency metrics	Build transit weighted network graph	Visualizations, report	Neighbourhood data analysis
Progress Report 2	March 10	Data Pipeline & Metric validation	Centrality, clustering	Dashboard prototype	Coverage analysis
Final Report	April 6	Technical review, submit	Methodology section	Editing, Final Visuals	Analysis section
Presentation	April 7/9	Accuracy review	Analysis slides	Presentation design	Coverage slides

5. Acknowledgement

No assistance was received from the TA in preparing this proposal. We thank the City of Winnipeg Open Data Portal and Statistics Canada for providing publicly accessible datasets.

---

6. Consent Declaration

We, the undersigned, have reviewed this document and consent to submit it as the project proposal for our group project.

Member	Initials for Signature	Date
Ahmed Hasan	ASA	January 19th, 2026
Cathy Li	CL	January 19th, 2026
Stephenie Michael	SCM	January 19th, 2026
Sudipta Sarker	SS	January 19th, 2026

A. References & Datasets

- City of Winnipeg. "Winnipeg Transit GTFS Feed." Transit.land. https://www.transit.land/feeds/f-cbfg-winnipegtransit
- City of Winnipeg. "Recent Transit On-Time Performance Data." Winnipeg Open Data. https://data.winnipeg.ca/Transit/Recent-Transit-On-Time-Performance-Data/gp3k-am4u
- City of Winnipeg. "Transit Pass-ups." Winnipeg Open Data. https://data.winnipeg.ca/Transit/Transit-Pass-ups/mer2-irmb
- City of Winnipeg. "Estimated Daily Passenger Activity." Winnipeg Open Data. https://data.winnipeg.ca/Transit/Estimated-Daily-Passenger-Activity/bv6q-du26
- City of Winnipeg. "Map of Neighbourhoods." Winnipeg Open Data. https://data.winnipeg.ca/dataset/Map-of-Neighbourhoods/iw67-pm65
- City of Winnipeg. "Community Characterization Area." Winnipeg Open Data. https://data.winnipeg.ca/City-Planning/Community-Characterization-Area/d9gq-xua6
- Statistics Canada. "2021 Census of Population." https://www12.statcan.gc.ca/census-recensement/2021/geo/sip-pis/index-eng.cfm
- Hagberg, A., Schult, D., & Swart, P. "NetworkX: Network Analysis in Python." https://networkx.org/
- Jordahl, K. et al. "GeoPandas: Python Tools for Geographic Data." https://geopandas.org/
- Filipe. "Folium: Python Data Visualization Library." https://python-visualization.github.io/folium/