

그래프 분할을 이용한 서울 수도권 지하철역들의 분류

박종수* · 이금숙**

요약 : 수도권 지하철 시스템은 지하철 역을 결절점으로 하고 이들을 연결하는 선로를 간선으로 구성된 그래프로 표현할 수 있다. 본 논문에서 서울 수도권 지하철 시스템의 통행흐름을 바탕으로 그래프를 거의 비슷한 그룹들로 분할하여 지하철역들의 분류와 지하철 승객들의 통행 특성을 연구한다. 그래프의 각 간선을 통과하는 승객수를 교통카드 트랜잭션 데이터베이스에서 추출하여 그 간선의 가중치로 둔다. 그래프 분할 문제는 NP-완전 문제에 속하기 때문에, 본 논문에서 지하철 시스템의 그래프를 분할하기위하여 휴리스틱 알고리즘을 제안한다. 그 휴리스틱 알고리즘은 두 개의 선택 가능한 목적함수들 중에서 하나를 사용하는 데, 첫 번째 목적함수는 다른 그룹들에 속한 결절점들을 연결하는 간선들의 가중치의 합을 최소화하는 것이고 두 번째는 전체 지하철 승객들에 대해 승차역과 하차역이 같은 그룹에 속한 승객들의 비율을 최대화하는 것이다. 실험결과에서 각 그룹에 속한 지하철역들과 간선들을 색깔로 구분하여 지도상에 표시하고 그룹별 기종점 행렬로 지하철 승객들의 통행 특성을 분석한다.

주요어 : 서울 수도권 지하철 시스템, 교통카드 트랜잭션 데이터베이스, 지하철 승객의 통행 특성, 그래프 분할, 기종점 행렬

1. 서론

도시민의 통행이 집중되는 교통 결절지나 중심 지점들의 공간적 배열과 이들 간의 통행흐름을 바탕으로 기능적 지역그룹을 파악하는 것은 지리연구의 핵심내용 중 하나이다. 또한 이는 교통정책 수립은 물론 질병의 확산 등 다양한 역학적 문제에의 적용까지 실용적인 적용이 많아 도시정책 및 계획 수립에서도 관심이 모아지는 부분이므로 이를 위한 다양한 연구가 시도되고 있다(Lee, *et al.*, 2012; Roth, *et al.*, 2011; 이금숙 외, 2010; Chen *et*

al., 2009; Guo, 2009; Assuncao, *et al.*, 2006).

본 논문에서는 그래프분할방법을 적용하여 우리나라 수도권 지하철을 사용하는 승객들의 통행 특성을 반영하여 지하철 시스템의 역들의 그룹을 분류하고 그의 특성을 파악하고자 한다. 그래프 분할 방법에 의하여 얻어지는 지하철역 분할 그래프는 각 지하철역의 분류 표시와 전체적인 분포를 한 번에 파악할 수 있게 하며, 보다 세분화된 분할을 적용하면 특정 지역의 지하철 승객들과 다른 지역의 승객들의 통행 특성 비교를 쉽게 할 수 있을 것이다.

그래프 분할문제는 여러 개의 결절점들과 그 결

이 논문은 2012년도 성신여자대학교 학술연구조성비 지원에 의하여 연구되었음.

* 성신여자대학교 IT학부 교수

** 성신여자대학교 지리학과 교수

절점들을 연결하는 연결선으로 구성된 그래프에서 절절점들을 몇 개의 그룹으로 분할하는 문제로 그룹들 사이에 걸쳐진 연결선 수가 최소가 되는 분할을 찾는 문제에서 출발하였으나 연결선에 가중치를 주어 가중치의 합이 최소가 되는 분할을 찾는 등 다양한 변형이 시도되고 있다. 이러한 그래프 분할 문제는 조합 최적화를 찾는 NP-완전 문제이므로 목적함수에 최적의 해답을 얻기가 어렵다. 본 논문에서는 널리 알려진 그래프 분할 프로그램을 이용하여 서울 수도권 지하철 시스템을 나타내는 그래프를 분할하여 지하철역들을 분류하는 휴리스틱 알고리즘을 제시하였다.

서울 수도권 지하철 승객들의 승차와 하차에 대한 정보를 담은 교통카드 트랜잭션 데이터베이스로부터 다양한 종류의 패턴을 추출해내는 연구가 이루어져오고 있다(Lee, *et al.*, 2012; Goh, *et al.*, 2012; Lee, *et al.*, 2011; 박종수·이금숙, 2011; 이금숙 외, 2010; 박종수 외, 2010; 박종수·이금숙, 2008; 박종수·이금숙, 2007; 이금숙·박종수, 2006). 교통카드 트랜잭션 데이터베이스는 하루 약 1,000만 이상의 트랜잭션들로 구성되고, 각 트랜잭션은 한 승객이 버스나 지하철을 한 번 승차하고 하차한 것에 대한 정보를 포함하고 있다. 트랜잭션에서 얻을 수 있는 정보는 승차장소, 승차 시간, 하차장소, 하차 시간, 환승횟수 등이 있다. 실제 교통카드 데이터베이스에서 어떤 종류의 지식이나 패턴을 탐사해내려면 교통 네트워크의 분석과 승객들의 통행 특성 분석이 필요하다. 서울 수도권 지하철 시스템을 절절점과 간선으로 구성된 그래프로 나타내면 절절점은 지하철역을 표현하고 간선은 두 지하철역을 연결하는 것을 표현한다. 이 그래프에서 간선의 가중치는 단순 연결 여부 또는 통과 승객수로 나타낼 수 있다.

본 논문에서는 지하철 시스템의 그래프를 p 개의 거의 비슷한 그룹들로 분할하여 지하철역들을 분류하고 지하철 승객들의 통행 특성을 파악하는 방법에 대해서 서술한다. 서울 수도권 지하철 네트

워크와 승객들의 흐름을 표현하는 그래프는 세 가지 종류로 나눌 수 있다. 첫 번째 그래프는 절절점과 간선으로 구성된 그래프의 간선의 가중치가 1로 나타내어 단순히 절절점과 절절점의 연결 유무만 표시하는 단순 그래프(simple graph)이다. 두 번째 그래프는 간선의 가중치가 그 간선을 통과하는 승객수로 나타내는 가중치 그래프(weighted graph)이다. 세 번째 그래프는 승객의 교통카드 트랜잭션에서 승차역과 하차역만 추출해내어 두 지하철역 사이를 직접 연결하는 간선을 만들어 통과 승객수를 가중치로 두는 OD-flow 그래프이다. 원래 컴퓨터 과학의 그래프 이론 분야에서 연구되는 그래프 분할(graph partitioning) 문제는 NP-complete 문제(Delling, *et al.*, 2010; Karypis and Kumar, 1988)로 많은 휴리스틱 알고리즘이 연구되고 있다. 본 논문에서는 많이 알려진 그래프 분할 프로그램인 METIS 프로그램(Karypis, 2011; Karypis and Kumar, 1988)을 이용하여 서울 수도권 지하철 시스템의 그래프를 분할하는 휴리스틱 알고리즘을 제안하고 실험 결과를 분석한다. 제안된 휴리스틱 알고리즘에 사용되는 목적함수는 분할된 그룹들을 연결하는 간선들의 가중치의 합을 최소화하는 것과 전체 승객들 중에서 동일 그룹에 속한 지하철역에서 출발하고 도착하는 승객들의 퍼센티지를 최대화하는 것으로 설정된다. 실험 결과 분석에서 분할된 각 그룹을 서로 다른 색깔로 그려서 지하철 승객들이 이용하는 그룹별 지하철역들의 분포를 파악하게 하고 지하철 승객들의 그룹별 기종점 행렬을 계산하여 승객들의 통행 특성을 분석하도록 한다.

논문의 구성을 살펴보면, 2절에서 서울 수도권 지하철 네트워크를 절절점과 간선을 가지는 그래프로 표현하는 방법과 각 간선의 가중치를 계산하는 방법을 설명한다. 그리고 이 그래프들을 분할할 때 적용하는 목적함수를 정의한다. 3절에서 그래프 종류와 목적함수에 따라 지하철역들을 분할하는 알고리즘을 제안하고 단계별로 그 내용을 기

술한다. 4절에서 서울 수도권 지하철 시스템의 교통카드 트랜잭션 데이터베이스를 제안된 알고리즘에 적용하여 세 종류의 그래프들을 분할한 결과를 승객들의 통행 특성에 맞추어 상세히 분석한다. 5절에서 이 논문의 결론을 맺는다.

2. 서울 수도권 지하철 시스템의 그래프 표현과 목적 함수

1) 서울 수도권 지하철 시스템의 그래프 표현

서울 수도권 지하철 시스템의 지하철역을 결절점(node or vertex)으로 표현하고 두 지하철역들이 연결되어 있는 정보는 간선(edge)으로 표현하면 지하철 시스템은 결절점들과 간선들로 구성된 그래프로 나타낼 수 있다. 그림 1은 서울 수도권 지하철 노선도(서울 메트로 노선도 서비스, 2012)의 일부분을 나타내고 있다. 이 그림에서 각 결절점은 특정 지하철역을 나타내고 간선은 한 지하철역과 다른 지하철역 사이의 연결을 표현하고 있다. 그림에서 “충무로”역은 3호선과 4호선이 겹치는 환승역으로 내부가 빈 원으로 표시하였다. “명동”역과 같이 환승역이 아닌 역은 작은 점으로 표시하고 있다.

그림 1의 표현 방법을 더 단순화하면 결절점들



그림 1. 서울 수도권 지하철 노선도 서비스의 일부분

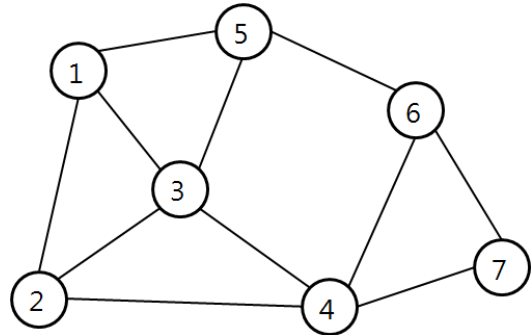


그림 2. Simple Graph

과 간선들로만 간결하게 서울 수도권 지하철 시스템을 하나의 그래프로 표현해 낼 수 있다. 설명을 위한 그림으로 7개의 결절점들과 간선들로 구성된 단순 그래프가 그림 2에 그려져 있다. 이 그림에서 각 번호를 가지고 있고, 이 번호는 지하철 시스템에서 각 역에 대한 정보인 역 이름, 노선 번호, 위도, 경도 등을 가지고 있다. 각 간선은 결절점과 결절점의 연결 여부를 나타내고 있다. 그림 2의 간선은 두 결절점들의 연결 여부만 나타내고 있어서 각 간선의 가중치는 1로 표시한다. 가중치 그래프는 각 간선에 가중치를 주어서 간선의 중요도나 흐름의 양을 표현해내는 데 그림 3과 같은 그래프이다. 이 그래프의 간선의 가중치는 각 간선을 통과하는 승객수를 나타낸다. 그림 2에서 결절점 5와 결절점 3은 연결되어 있다는 것만 나타내지만, 그림 3에서 두 결절점들 사이에는 가중치가 2가 있어서 두 결절점들 사이에 승객 2명이 통과하였다는 것을 설명한다. 그림 4는 결절점 7개를 가진 완전 그래프를 보여주고 있다. 그림 4의 각 간선은 지하철 시스템의 그래프에서는 승객이 승차한 지하철역(Origin)과 하차한 지하철역(Destination)을 연결하는 정보를 나타내고 그 가중치는 두 지하철역들 사이에 오고 간 승객들의 수를 나타내는 그래프이므로 OD-flow graph로 부른다. 서울 수도권 지하철 시스템을 표현하는 그림 3과 4의 차이점은 승객의 이동을 어떻게 표현하는가에 달려

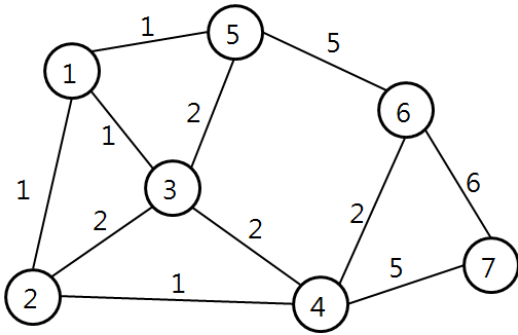


그림 3. Weighted Graph

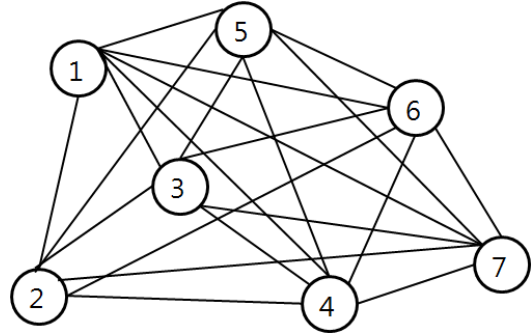


그림 4. OD-flow Graph

다. 한 승객이 결절점 1에서 출발하여 결절점 7에 도착하였다면, 그림 3의 표현 방법에 의하면 먼저 두 결절점 사이에 최단 경로를 정하고 이 경로 상의 각 간선에 가중치를 1씩 더한다. 예를 들면, 결절점 1과 7 사이의 최단 경로가 <1, 3, 4, 7>이라면 이 경로상의 간선들 (1, 3), (3, 4), (4, 7)에 가중치를 1씩 더한다. 그렇지만, 그림 4의 표현에 따르면 결절점 1과 7 사이의 간선 (1, 7)에 가중치 1을 더한다.

본 논문에서는 다음과 같은 세 종류의 지하철 시스템의 그래프를 사용한다:

a) Simple Graph(단순 그래프): 그림 2와 같이 각 지하철역을 결절점으로 나타내고 단순 간선으로만 그래프를 표현한다. 이 경우는 수도권 지하철 시스템의 위상학적인 특성만 고려하는 것으로 어떤 지하철역이 다른 지하철역과 연결되어 있는지를 나타내고 있다.

b) Weighted Graph(가중치 그래프): 그림 3과 같이 각 간선의 가중치는 그 간선을 통과하는 지하철 승객들의 숫자로 표현한다. 지하철 승객이 승차역에서 지하철을 타고 하차역에서 내리면, 승객이 이동하는 최단 경로를 구하여 경로상의 간선에 승객수를 더하여 가중치를 얻는다. 그러면 각 간선의 가중치는 통과한 승객수가 되어서 그래프를 분할하면 승객들의 그룹별 통행특성을 알 수 있다.

c) OD-flow Graph: 그림 4와 같은 그래프는 승객들의 통행패턴이 출발-도착(Origin-Destination) 개념으로 분석되어지는 것으로 승객들이 이동하는 중간 통과역에 대한 정보를 사용하지 않고 승차역과 하차역의 정보만 이용하여 통행 특성을 분석해낸다.

2) 그래프 분할과 목적 함수

그래프 분할(graph partition)은 컴퓨터 과학에서 많은 응용 분야를 가진 아주 중요한 문제다. 그 응용 분야로는 VLSI 설계, 컴퓨터 비전, 영상 분석, 분산 컴퓨팅, 노선 기획(route planning) 등을 포함한다. 그래프 분할 문제는 주어진 그래프의 결절점들을 p 개의 그룹들로 분할하는 데 다음과 같은 두 조건을 만족해야 한다: 첫째로 각 그룹에 속한 결절점들의 개수는 거의 같아야 하고, 둘째로 그래프 분할에서 edge-cut이라 부르는 분할된 그룹들을 연결하는 간선들의 가중치의 합이 최소가 되어야 한다는 조건이다(Karypis and Kumar, 1998). 그래프 분할 문제는 NP-complete 문제에 속하여 최적의 해법을 찾아내는 알고리즘은 아직 발표 되지 않고 있다. 널리 알려진 METIS 프로그램(Karypis, 2011; Karypis and Kumar, 1988)은 주어진 그래프를 입력 파라미터와 옵션에 따라 무작위로 초기 분할을 하여 원하는 개수의 그룹들로

분할하는 프로그램이다. 본 논문에서는 그래프를 분할할 때 목적함수를 정의하고 이것에 적합한 그래프 분할을 얻기 위하여 METIS 프로그램을 여러 번 반복 실행하여 목적함수를 만족하는 최선의 결과를 구한다.

서울 수도권 지하철 시스템의 그래프를 분할할 때 앞에서 제시한 조건들을 적용하면, 분할된 각 그룹에 소속된 지하철역들의 개수는 거의 같고 인접해야 한다는 것이고 두 번째로 분할된 그룹들 사이에 연결된 간선들의 가중치의 합이 최소가 되어야 한다. 본 논문에서는 edge-cut의 값을 분할된 그룹들을 연결하는 간선들의 가중치의 합으로 둔다. 두 번째 조건에서 그룹들로 나누어지게 하는 간선들의 가중치 값이 최소인 것을 minimum cut이라 한다. 먼저 간단한 그래프에서 그룹들로 분할하는 것을 살펴보자. 그림 2에서 7개의 결절점을 두 개의 그룹으로 나눌 때 minimum cut이 되는 결과는 그림 5와 같이 표현할 수 있다. 여기서 edge-cut에 해당되는 간선들은 (2, 4), (3, 4), (5, 6)이고 edge-cut의 값은 간선들의 개수에 해당되는 3이다. 그림 2의 그래프에서 edge-cut의 값이 3보다 작은 값으로 거의 균등한 개수의 결절점들을 갖는 그래프 분할은 없다. 그림 3은 가중치가 있는 간선들로 이루어진 그래프로 두 개의 그룹으로 그래프를 분할하고자 하면 edge-cut의 값이 최소가 되는 분할이어야 한다. 그림 5와 같은 edge-cut을

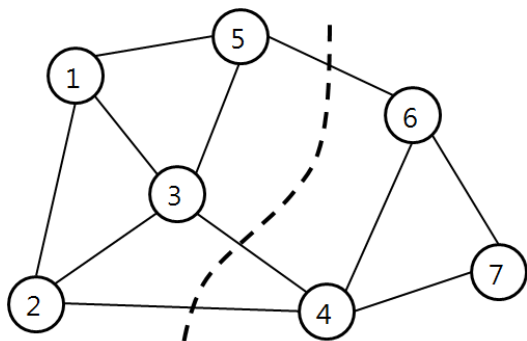


그림 5. 분할된 단순 그래프

가지게 되면 그림 3에서 이 edge-cut의 값은 가중치의 합으로 8이 되고, 그림 6과 같은 edge-cut인 (1, 5), (2, 4), (3, 4), (3, 5)의 가중치의 합은 6이 된다. 그림 3의 그래프에서 두 그룹으로 나누는 모든 경우의 분할에서 edge-cut의 가중치의 합이 최소가 되는 값은 6이다. 그러므로 그림 3의 그래프 분할의 결과는 그림 6과 같은 두 그룹의 분할로 나타난다. 본 논문에서는 그림 5와 6에서 설명한 것과 같이 그래프 분할 문제의 첫 번째 목적함수로 edge-cut의 값을 최소화시키는 minimum cut으로 둔다.

그림 4와 같은 OD-flow를 나타내는 그래프의 분할에 대한 평가는 minimum cut과 같은 기준을 적용하기에는 조금 모호한 면이 있다. 이를 보완하기 위하여 다른 방식을 살펴보면, 데이터 마이닝에서 군집화(clustering)의 유효성 평가를 위한 척도(measure)로 두 가지 접근 방식을 고려하고 있다(Tan *et al.*, 2006). 첫 번째 기법은 엔트로피(entropy), 순도(purity), F-measure와 같은 분류화 척도들을 사용한다. 이 척도들은 분할된 각 부분이 단일 분류의 객체들을 어느 정도 가지고 있는지를 평가한다. 두 번째 기법은 자카드(Jaccard) 척도와 같이 같은 분류에 속한 두 개의 객체가 어느 정도로 같은 군집에 존재하는지를 측정하기 위한 방법이다. 여러 척도들 중에서 지하철 승객들의 통행 특성을 분석하기 위해 순도(purity)의 개념을

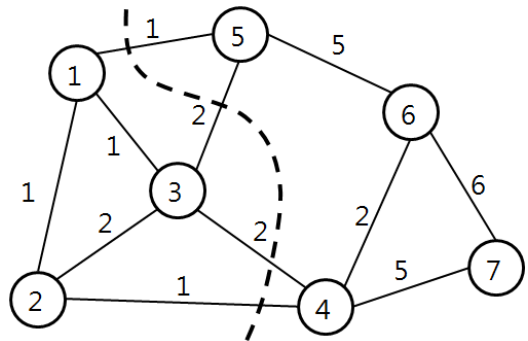


그림 6. 분할된 가중치 그래프

적용한다. 서울 수도권 지하철 시스템의 그래프 분할에서 순도는 같은 그룹에 속한 승차역과 하차역을 이용하는 승객들이 전체 지하철 승객들에 대해서 어느 정도의 비율을 가지는 지를 측정하는 것이다. 본 논문에서 지하철 시스템의 그래프 분할 문제의 두 번째 목적함수로 순도를 최대화시키는 최대 순도(maximum purity)로 둔다.

서울 수도권 지하철 시스템의 그래프를 분할할 때 적용할 목적함수로 minimum cut과 maximum purity를 채택하였으므로 그래프 종류에 따른 목적함수의 적용은 표 1에 설명한다. 표 1에서 그래프 종류와 목적함수에 따라 구분되는 유형을 나타내고, T1은 Type 1을 표시한다. 목적함수를 간단히 설명하면 다음과 같다.

a) Minimum cut: 그래프를 p 개의 그룹들로 분할할 때 그룹들을 연결하는 간선들의 가중치의 합인 edge-cut의 값이 최소가 되는 분할을 구한다.

b) Maximum purity: 그래프 분할에서 전체 승객들에 대하여 승차역과 하차역이 같은 그룹에 속한 승객들의 퍼센티지(purity)를 구하여 이것을 최대화하는 분할을 구한다.

표 1. 그래프 분할에 적용되는 목적 함수

그래프 \ 목적함수	minimum cut	maximum purity
Simple Graph(그림 2)	T1	-
Weighted Graph(그림 3)	T2	T3
OD-flow Graph(그림 4)	T4	T5

3. 지하철 시스템의 그래프 분할 알고리즘

서울 수도권 지하철 시스템의 그래프를 분할하는 프로그램은 표 1에 따라 5가지의 유형들을 찾아 내는 알고리즘들을 가져야 한다. 유형 1(T1)을 구하는 알고리즘은 그림 7에서 주요 단계들을 기술하고, 유형 2와 3에 대한 그래프 분할 알고리즘은 그림 8에서 기술한다. 유형 4와 5에 대한 그래프 분할 알고리즘은 그림 9에서 기술한다.

그림 7은 서울 수도권 지하철 시스템을 단순 그래프로 표현하여 METIS 프로그램으로 minimum cut을 찾아서 지하철역과 연결 간선을 그래픽으로 보여줄 수 있는 파일을 만들어 내는 과정을 설명한 것이다. 단계 3에서 METIS 프로그램을 실행하는 여러 option 중에서 -seed의 숫자를 변경하면 그래프의 초기 분할의 모습이 달라져서 minimum cut으로 구한 p 개로 분할된 그룹들이 조금씩 달라지고 이에 따라 edge-cut의 값도 차이가 난다. seed의 값을 변경하여 n 번의 반복 실행으로 그 중에서 가장 작은 edge-cut의 값을 갖는 seed 번호를 찾아서 minimum cut에 해당되는 분할을 구하게 된다. 이런 반복 실행을 하는 이유는 그래프 분할 문제가 NP-완전 문제에 속해 최적의 해결을 찾아내지 못하므로 여러 번 실행하여 그 중에서 최소값을 얻기 위한 것이다. 실험에서 n 은 20만으로 설정하였다. 단계 4에서 분할된 그룹을 색깔로 구분하여

```

Void SimpleGraphMinimumCut(int p, int n)
{
    단계 1: 통합된 지하철역, 노선 연결, 환승역에 대한 자료를 입력받는다.
    단계 2: 지하철역을 결절점으로 하고 노선 연결성을 간선으로 하여 METIS 프로그램의 입력용 파일을 만든다.
    단계 3: METIS 프로그램의 옵션을 변경하여 n번 실행하여 minimum cut을 갖는 그래프 분할 결과인 p개의 그룹들을 구한다.
    단계 4: 분할된 그룹을 색깔로 구분하여 결절점과 간선을 그래픽으로 그린다.
}
    
```

그림 7. Simple Graph에서 minimum cut을 찾는 알고리즘

```

Void WeightedGraphMinimumCutOrMaximumPurity(int p, int n)
{
    단계 1: 통합된 지하철역, 노선 연결, 환승역에 대한 자료를 입력받는다.
    단계 2: 분리된 지하철역과 간선에 대한 자료를 입력받는다.
    단계 3: 각 지하철 승객의 트랜잭션을 읽어서 승차역과 하차역을 찾아내고 최단
            경로를 구한다. 최단 경로를 구성하는 각 간선의 가중치를 1씩 증가시킨
            다.
    단계 4: 지하철역을 결절점으로 하고 노선 연결성을 간선으로 하여 간선의 가중치
            를 적용하는 METIS 프로그램의 입력용 파일을 만든다.
    단계 5: 만약 목적 함수가 minimum cut이면, METIS 프로그램의 옵션을 변경하여
            n번을 실행하여 minimum cut을 갖는 분할된 p개의 그룹들을 구한다.
    단계 6: 만약 목적 함수가 maximum purity이면, METIS 프로그램의 옵션을 변경
            하여 n번 실행하여 최대 순도를 갖는 분할된 p개의 그룹들을 구한다.
    단계 7: 분할된 그룹을 색깔로 구분하여 결절점과 간선을 그래픽으로 그린다.
}
    
```

그림 8. Weighted Graph에서 minimum cut 또는 maximum purity를 찾는 알고리즘

각 그룹에 속한 지하철역들은 같은 색깔로 그려서 전체적인 분포를 쉽게 파악하도록 한다. 그 결과의 예제로 그림 10에 지하철역과 간선들을 분할에 따른 색깔로 그려진 것을 보여주고 있다.

그림 8은 가중치 그래프로 표현된 지하철 시스템의 그래프를 METIS 프로그램을 이용하여 minimum cut 또는 maximum purity를 갖는 분할된 그룹들을 구하는 단계를 설명하고 있다. 단계 3에서 지하철 승객의 승차역과 하차역 사이의 최단 경로를 구하는 과정에서 환승역에서 환승하는 시간을 고려하여 최단 경로를 구해야 하므로 단계 1과 2에서 통합된 지하철 시스템과 분리된 지하철 시스템의 정보를 다르게 받아서 처리한다. 단계 4에서 METIS 프로그램의 입력 파일에서의 결절점과 간선의 정보는 통합된 지하철 시스템의 결절점과 간선을 사용한다. 예를 들면, 그림 1에서 “종로3가”역은 통합된 시스템에서는 하나의 역으로 간주되지만 분리된 시스템에서는 3개의 노선이 지나가므로 “1호선 종로3가”역, “3호선 종로3가”역, “5호선 종로3가”역으로 분리되어 최단 경로를 구할 때 이용된다. 단계 5와 6에서는 목적 함수에 따라 얻는 그룹들이 다르게 계산되는 것을 보여준다.

그림 9는 OD-flow graph를 분할하는 알고리즘을 기술하고 있다. 단계 1에서 지하철역들로 구성된 결절점들로부터 먼저 OD(Origin-Destination) matrix를 만들고, 단계 2에서는 분리된 지하철역과 간선에 대한 자료를 입력받는다. 단계 3에서 입력되는 지하철 승객 트랜잭션의 승차역과 하차역의 정보는 분리된 지하철역 이름을 사용하므로 이것에 대응되는 통합된 지하철역 이름을 찾아내어 승차역(Origin)과 하차역(Destination)의 정보로 이용한다. 각 지하철 승객의 트랜잭션의 승차역과 하차역에 해당되는 OD matrix의 값을 증가시킨다. 단계 4에서 OD matrix를 간선의 가중치를 갖는 그래프로 변환하여 METIS 프로그램의 입력으로 사용하게 한다. 단계 6에서 METIS 프로그램으로 분할된 그룹들의 purity를 계산하는 방법은 먼저 GroupOD[1:p][1:p]를 구한다. 한 승객의 트랜잭션에서 승차역이 그룹 i에 속하고 하차역이 그룹 j에 속하면 GroupOD[i][j]의 값이 1 증가한다. 단계 6의 그룹 정보와 단계 3의 OD matrix 정보를 이용하여 GroupOD matrix의 값을 계산하면, purity의 값은 다음 식으로 구한다:

```

Void ODflowGraphMinimumCutOrMaximumPurity(int p, int n)
{
    단계 1: 통합된 지하철역, 노선 연결, 환승역에 대한 자료를 입력받는다. 이 자료
        에 따라 지하철역들의 OD matrix를 만든다.
    단계 2: 분리된 지하철역과 간선에 대한 자료를 입력받는다.
    단계 3: 각 지하철 승객의 트랜잭션을 읽어서 승차역과 하차역을 찾아내어 OD
        matrix의 해당 cell의 값을 1씩 증가시킨다.
    단계 4: OD matrix에 대응되는 그래프를 METIS 프로그램의 입력용 파일을 만든
        다.
    단계 5: 만약 목적 함수가 minimum cut이면, METIS 프로그램의 옵션을 변경하여
        n번을 실행하여 minimum cut을 갖는 분할된 p개의 그룹들을 구한다.
    단계 6: 만약 목적 함수가 maximum purity이면, METIS 프로그램의 옵션을 변경
        하여 n번 실행하여 최대 순도를 갖는 분할된 p개의 그룹들을 구한다.
    단계 7: 분할된 그룹을 색깔로 구분하여 결절점과 간선을 그래픽으로 그린다.
}
    
```

그림 9. OD-flow Graph에서 minimum cut또는 maximum purity를 찾는 알고리즘

$$\text{purity} = \frac{\sum_{i=1}^p \text{GroupOD}[i][i]}{\sum_{i=1}^p \sum_{j=1}^p \text{GroupOD}[i][j]}$$

단계 6에서 n번의 그래프 분할에서 구한 purity 중에서 최대값을 갖는 그래프 분할 결과를 저장하여 최종 그룹들로 정한다. 그림 7과 8의 알고리즘에서도 후속 처리를 통해 Group OD matrix를 계산해내어 지하철 승객들의 통행 특성 분석에 사용한다.

4. 분할된 그래프의 분석

이 절은 앞에서 제안한 알고리즘을 서울 수도권 지하철 시스템의 그래프에 적용하여 얻은 실험 결과에 대하여 상세히 서술한다. 실험에 사용된 컴퓨터의 운영체제는 MS Windows 7 64bit이고, 프로그램 개발 환경으로 MS Visual Studio 2008 C++언어를 사용하였다. 사용된 컴퓨터 하드웨어의 주요 사양으로 CPU가 Intel i920 2.67GHz 이고 메인 메모리는 16GB DDR3 RAM이다. 제안된 알고리즘의 입력으로 사용된 교통카드 트랜

잭션 데이터베이스는 2007년 5월의 어느 하루 동안 수도권 대중교통을 사용한 승객들의 트랜잭션들이다. 전체 트랜잭션들의 개수는 10,856,337 이고, 이 중에서 지하철 승객들의 트랜잭션 수는 5,174,181이다. 이 날 기준의 서울 수도권 지하철 시스템에서 지하철역들의 개수는 378이다.

그림 10, 11, 12는 수도권 지하철역들을 2개 그룹으로 분류할 때 표 1에서 설명한 그래프의 종류와 목적 함수에 따라 얻어진 결과를 보여주고 있다. 그림 7의 알고리즘으로 구한 그림 10의 P2T1에서 P2는 p=2인 경우로 그래프를 2 그룹으로 분류한다는 의미이고 T1은 표 1에서 simple graph에서 minimum cut을 목적함수로 하는 유형 1을 의미한다. 실험에서 n=20만으로 두어, 20만 번의 METIS 프로그램을 수행하여 목적함수에 맞는 결과를 얻는다. 그림 10의 점은 지하철역을 표시하고 점사이의 선분은 두 지하철역들 사이에 연결된 지하철 선로를 표시한다. 지하철역에 해당되는 점의 위치는 위도와 경도에 따라 그려진 것이다. 그림 10에서 그래프 분할 결과는 지하철역들이 남북으로 두 그룹으로 나누어진 모습을 보여주고 있다. 그림 10의 P2T1 그래프의 후속 처리를 통하여 지하철 승객들의 그룹별 이동 상황을 표 2에서 설

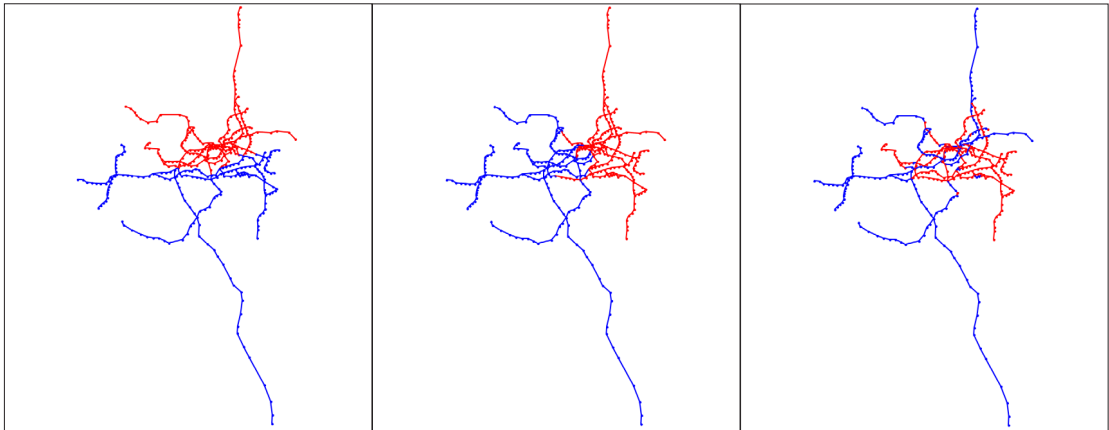


그림 10. P2T1

그림 11. P2T3

그림 12. P2T4(P2T5)

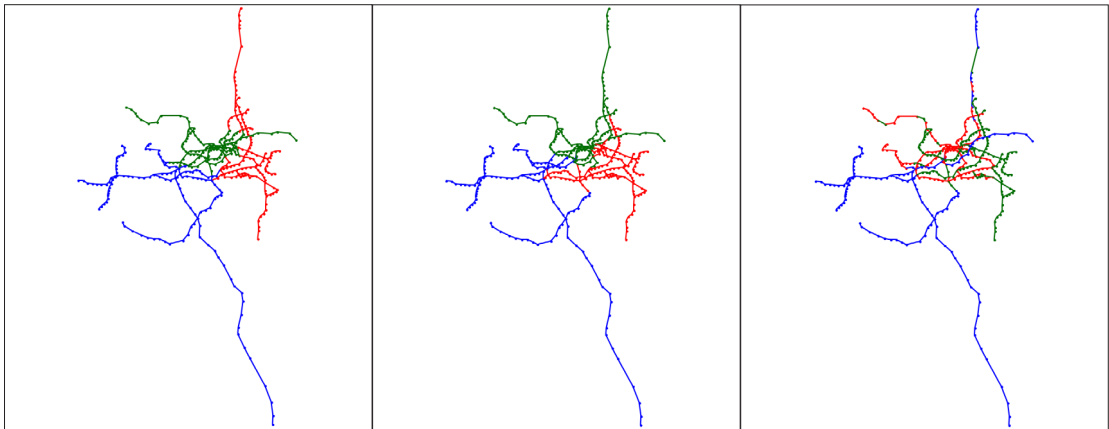


그림 13. P3T1

그림 14. P3T3

그림 15. P3T4(T3T5)

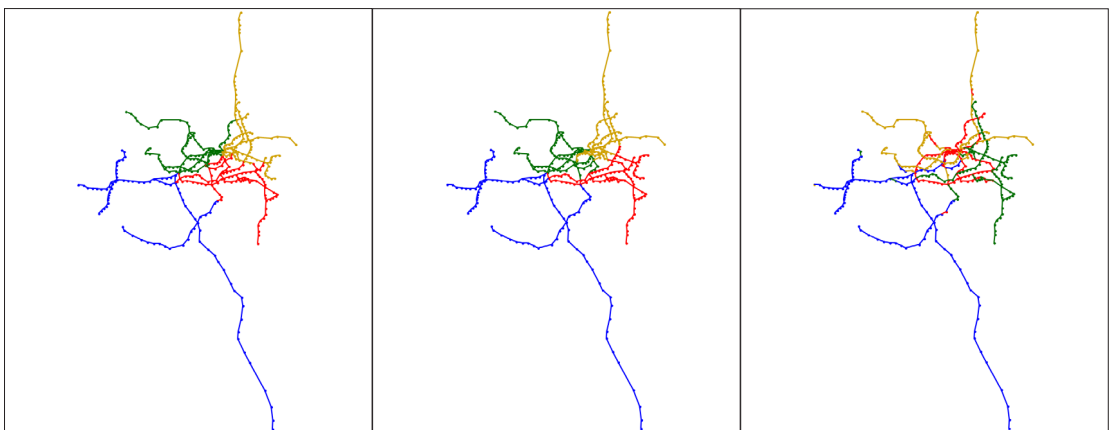


그림 16. P4T2

그림 17. P4T3

그림 18. P4T4(P4T5)

명한다.

표 2에서 청색 그룹과 적색 그룹의 분류 기호로 G_B 와 G_R 로 표기하고 각 그룹에서 승차하고 하차하는 승객수를 행렬로 설명하고 있다. 예를 들면, 그룹 G_B 에 속한 지하철역에서 승차하여 그룹 G_B 에 속한 지하철역에서 하차하는 승객수는 1,716,137명이다. 표에서 $|G_i|$ 는 그룹 G_i 에 속한 지하철역들의 개수를 나타낸다. 두 그룹에 속한 지하철역들의 차이는 4개로 거의 비슷한 개수로 그래프가 분할되었음을 보여준다. 표 2에서 edge-cut의 값 7은 지하철 시스템의 단순 그래프가 두 개의 그룹들로 분할할 때 7 개의 간선들이 분할된 두 그룹들 사이에 있다는 것을 의미한다. 표 2에서 purity는 $(1716137+1763016)/5174181=0.6724$ 로 계산되어 표기된다. 여기서 전체 지하철 승객수는 5,174,181명이다. 결과 그림들에서 보여주지 못한 P2T2의 결과 그래프는 그림 10의 P2T1과 거의 같지만 5호선 동쪽의 지하철역들이 북쪽 그룹인 적색 그룹에 속하는 차이점이 있다. P2T2에 대한 표 3은 통과 승객수로 간선의 가중치로 이루어진 가중치 그래프를 minimum cut으로 분할하였을 때 얻어진 자료이다. 표 3에서 edge-cut의 값은 분할된 두 그룹들을 연결하는 간선들의 가중치의 합을 나타내므로 표에서 간단히 생각하면 edge-cut의 값은 두 그룹을 교차하는 승객들의 수

$809172+819341=1628513$ 으로 생각할 수 있다. 그러나 한 승객이 승차역에서 하차역으로 가는 최단 경로 상에서 승차역이 청색 그룹에 속하고 중간에 적색 그룹에 속한 지하철역을 통과하고 마지막으로 청색 그룹에 속한 하차역에 도달할 수 있다. 이런 경우에는 한 승객이 edge-cut에 속한 간선을 두 번 통과하므로 edge-cut의 값은 2가 더해진다. 그러므로 표 3에서 edge-cut의 값이 1,628,513 보다 큰 1,686,819가 된다. 표 4의 edge-cut의 값도 같은 성격을 가지고, 반면에 표 5의 edge-cut의 값은 OD-flow 그래프에서 구한 값으로 G_B 에서 G_R 로 가는 승객수와 G_R 에서 G_B 로 가는 승객수의 합과 정확히 일치한다.

그림 11은 간선의 가중치가 그 선로를 통과하는 승객수로 나타내는 weighted graph를 그림 8의 알고리즘에서 maximum purity로 두 그룹으로 분할한 결과 그래프 P2T3이고, 지하철역들이 동쪽 그룹과 서쪽 그룹으로 분할되는 것을 보여준다. 그림 11의 그래프에서 계산되어진 purity의 값은 표 4에서 0.6946이다. 이것은 약 69%의 승객들이 동일 그룹에 속한 지하철역에서 승차하고 하차한다는 것을 의미한다. 서울 수도권 지하철 시스템을 이용하는 승객들의 통행 특성은 동쪽의 승객들은 동쪽에 속한 지하철역들을 이용하고 서쪽의 승객들은 주로 서쪽의 지하철역들에서 승차와 하차를

표 2. P2T1(그림 10)의 Group OD Matrix

O \ D	G_B	G_R	$ G_i $
G_B	1,716,137	845,824	187
G_R	849,204	1,763,016	191

edge-cut=7, purity=0.6724

표 4. P2T3(그림 11)의 Group OD Matrix

O \ D	G_B	G_R	$ G_i $
G_B	1,335,128	799,034	184
G_R	781,165	2,258,854	194

edge-cut=2,087,047, purity=0.6946

표 3. P2T2의 Group OD Matrix

O \ D	G_B	G_R	$ G_i $
G_B	1,767,101	809,172	184
G_R	819,341	1,778,567	194

edge-cut=1,686,819, purity=0.6853

표 5. P2T4(그림 12)의 Group OD Matrix

O \ D	G_B	G_R	$ G_i $
G_B	1,001,651	683,293	184
G_R	661,134	2,828,103	194

edge-cut=1,344,427, purity=0.7402

하는 것을 알 수 있다. 나머지 31%의 승객들은 한 그룹의 지하철역에서 승차하여 다른 그룹의 지하철역에서 하차하고 있음을 결과로 보여준다. 2호선의 남쪽 부분에서 “봉천”역과 “신림”역까지 동쪽 그룹인 적색 그룹에 속하는 특징을 보여주고 있다. 표 3의 분할 목적 함수가 minimum cut인 경우와 표 4의 분할 목적 함수가 maximum purity인 경우를 비교하면 표 3에서 edge-cut의 값이 작음을 알 수 있고, 반대로 표 4의 purity가 표 3에 비해서 더 큰 것임을 보여주고 있다. 그림 9의 알고리즘으로 구한 그림 12는 지하철 승객의 승차역과 하차역 사이의 간선에 통과 승객수로 가중치를 둔 OD flow 그래프를 minimum cut이나 maximum purity의 목적 함수로 분할한 결과 그래프 P2T4와 P2T5인데 $n=20$ 만으로 주었을 때 같은 결과가 나옴을 보여주고 있다. 이 그래프의 분할은 공간적으로 인접한 간선들의 그래프를 분할하는 것이 아니고, 승객들의 승차역(Origin)과 하차역(Destination) 사이의 가상의 간선들의 OD-flow 그래프를 분할하는 것으로 동일 그룹에는 속하지만 인접하지 않은 지하철역도 있다. 예를 들면, “경마공원”역은 적색 그룹에 속하지만 그 역에 인접한 “선바위”역이나 “대공원”역은 다른 청색 그룹에 속한다. 이것은 적색 그룹에 속한 역들을 이용하는 승객들이 “경마공원”역을 많이 사용함을 설명하고 있다.

그림 13~15는 알고리즘에서 $p=3$ 인 경우로 수도권 지하철역들을 3개 그룹으로 분류할 때, 표 1의 T1-T5에 대한 분할된 그룹들을 나타내고 있다. P3T2의 결과 그래프는 그림 14의 P3T3 그래프와 유사하지만 7호선 “태릉입구”역에서 “노원”역까

지 녹색 그룹에 속하는 것과 중앙선의 “이촌”역과 “옥수”역이 적색 그룹에 속하는 것이 다른 점이다. 그림 13과 14를 비교해보면, 적색 그룹과 녹색 그룹에 속하는 지하철역들의 분포가 많이 다를 수 있다. 그림 13의 그래프는 지하철 선로의 연결 여부만 표시하는 simple graph의 분할 결과 그래프이고, 그림 14는 지하철 승객들의 승차역과 하차역의 최단 경로 상의 간선에 가중치를 준 weighted graph에서 목적 함수가 maximum purity에 의한 결과 그래프이다. 그림 14에 의하면, 1호선 북쪽 지역과 중앙선 지역의 지하철역들이 한 그룹인 녹색 그룹에 속한다는 것을 보여주고 있다. 그림 15는 P3T4와 P3T5의 결과 그래프가 동일한 내용이어서 한 그래프로 표시하였다. 그림 15에서 1호선의 “회룡”역과 “의정부”역이 주변 역들과 달리 2호선 지하철역들이 속한 적색 그룹에 속하는 것과 3호선의 북서쪽 역들도 적색 그룹에 속하는 것을 그림으로 보여주고 있다. 표 6은 그림 13의 P3T1 그래프에 관련된 자료이다. 분할된 3개의 그룹을 색깔별로 청색, 적색, 녹색 그룹으로 나타내고 기호로 G_B, G_R, G_G 로 표기한다. 표 6에서 edge-cut은 주어진 그래프를 3개의 그룹으로 분할할 때 잘려지는 간선의 개수로 12임을 보여주고 있다. 표 7은 그림 14에서는 표시하지 않았지만 통과 승객수가 가중치가 주어진 weighted graph를 minimum cut으로 분할하였을 때 얻어진 자료이다. 표 8과 9는 그림 14와 15에 대한 그룹 OD 행렬로 승객들의 이동 상황을 보여준다. 표 9에서 그룹 G_R 에 속한 지하철역에서 출발하여 그룹 G_R 에 속한 지하철역으로 하차하는 승객수가 아주 많음을 보여주고 이

표 6. P3T1(그림 13)의 Group OD Matrix

O \ D	G_B	G_R	G_G	Gi
G_B	839,916	317,665	408,573	127
G_R	302,562	999,045	515,551	125
G_G	386,893	509,432	894,544	126

edge-cut=12, purity=0,5283

표 7. P3T2의 Group OD Matrix

O \ D	G_B	G_R	G_G	Gi
G_B	843,428	296,067	315,896	122
G_R	284,773	1,164,440	479,037	127
G_G	305,234	489,941	995,365	129

edge-cut=2,483,770, purity=0,5804

표 8. P3T3(그림 14)의 Group OD Matrix

O \ D	G _B	G _R	G _G	G _i
G _B	807,655	258,171	310,742	122
G _R	248,090	1,263,106	491,015	127
G _G	298,729	488,260	1,008,413	129

edge-cut=2,583,174, purity=0.5951

에 따라 purity의 값이 다른 표들의 값보다 높음을 보여주고 있다.

그림 16~18은 알고리즘에서 $p=4$ 인 경우로 서울 수도권 지하철역들을 4개 그룹으로 분할할 때 분할 목적 함수에 따른 결과 그래프이다. 그림 16은 간선의 가중치가 그 간선을 통과하는 승객수로 나타나는 가중치 그래프에서 minimum cut이 되는 그룹들로 분할한 결과 그래프이다. 청색 그룹의 1호선 천안선과 인천선에 속한 지하철역들로 이루어진 그룹, 2호선 남부지역, 강남지역, 그리고 분당지역에 속한 지하철역들로 이루어진 적색 그룹, 서북쪽 지역의 지하철역들로 구성된 녹색 그룹, 마지막으로 북동지역의 지하철역들이 속한 황색 그룹으로 분할되고 있음을 알 수 있다. 결과 그림에 설명하지 않은 P4T1은 그림 16과 비슷하지만 5호선 동쪽의 지하철역들이 적색 그룹에 속하고 적색 그룹에 속한 중앙선과 6호선 지하철역들이 녹색 그룹에 속하는 것 등의 차이점이 있다. Maximum purity를 목적 함수로 한 그림 17의 P4T3의 결과 그래프에서 그림 16의 적색 그룹에 속한 4호선 지하철역들이 그림 17에서 각각 청색과 녹색 그룹에 속하는 것을 알 수 있다. 4호선을 이용하는 승객들은 2호선의 강남 지역에 하차역으로 하차하는 경우가 작음을 알 수 있다. 그림 15의 P3T4와 그림 18의 P4T4를 비교해보면 2호선을 중심으로 하는 적색 그룹이 P4T4에서 가운데로 축소된 모습을 보여주고 있다. 이런 특성은 T4의 목적 함수로 인하여 나타나는 데 T4는 승객들의 승차역과 하차역만 고려하는 OD-flow graph에서 minimum cut을 목적함수로 그래프를 분할하

표 9. P3T4(그림 15)의 Group OD Matrix

O \ D	G _B	G _R	G _G	G _i
G _B	657,929	376,713	84,991	125
G _R	367,307	2,163,746	467,328	129
G _G	81,355	484,590	490,222	124

edge-cut=1,862,284, purity=0.6401

는 것이다. 그림 18에서 1호선 북동지역의 “회룡”역과 4호선 남쪽지역의 “인덕원”역이 적색 그룹에 속하는 특이한 분류 특성을 보여주고 있다.

5. 결론

본 논문에서 서울 수도권 지하철 시스템을 절절점과 간선으로 구성된 그래프로 나타내고 이 그래프를 몇 개의 그룹으로 분할하는 방법을 연구하였다. 간선의 가중치를 통과하는 승객들의 수로 나타내는 그래프 표현 방법과 승객들의 승차역과 하차역에 간선을 두고 그 가중치로 통과 승객수로 한 OD-flow 그래프 표현 방법을 이용하였다. 그래프 분할 문제는 NP-완전 문제이므로 목적함수에 최적의 해답을 얻기가 어렵다. 본 논문에서는 널리 알려진 METIS 그래프 분할 프로그램을 이용하여 서울 수도권 지하철 시스템을 나타내는 그래프를 분할하는 휴리스틱 알고리즘을 제시하였다. 교통카드 트랜잭션 데이터베이스에서 지하철 승객들의 통행 과정을 추출하여 그래프의 간선의 가중치를 설정하였다. 실험 결과에서 지하철 시스템의 그래프를 노선의 연결성을 표시한 간선으로 구성된 단순 그래프로 표현하였을 때 2개의 그룹으로 분할하면 남쪽 그룹과 북쪽 그룹으로 나누어졌지만, 그래프의 간선의 가중치를 통과 승객수로 두고 전체 지하철 승객들에 대하여 같은 그룹의 승차역과 하차역을 사용하는 승객들의 퍼센티지를 최대화하는 목적함수에 의한 분할 결과 그래프

는 동쪽 그룹과 서쪽 그룹으로 나누어졌다.

본 논문의 실험 결과에서 지하철을 사용하는 승객들의 통행 특성을 반영하면 주어진 지하철 시스템의 분류 패턴도 그에 따른 특성을 가지고 있음을 보여주고 있다. 지하철 시스템을 나타내는 그래프를 3개의 그룹과 4개의 그룹으로 나눈 실험 결과를 색깔로 표시된 그래프로 보여주고 승객들의 통행 특성인 그룹별 기종점 행렬도 계산하여 그 특성을 상세히 분석하였다. 본 논문의 실험 결과인 지하철역 분할 그래프는 각 지하철역의 분류 표시와 전체적인 분포를 한 번에 파악할 수 있는 방법을 보여준다. 보다 세분화된 분할을 적용하면 특정 지역의 지하철 승객들과 다른 지역의 승객들의 통행 특성 비교를 쉽게 할 수 있을 것이다.

참고문헌

박종수·김호성·이금숙, 2010, “교통카드 트랜잭션 데이터베이스에서 지하철 탑승 패턴 분류,” 한국콘텐츠학회논문지 10(12), pp.91-100.

박종수·이금숙, 2007, “대용량 교통카드 트랜잭션 데이터베이스에서 통행패턴 탐사와 통행행태 분석,” 한국경제지리학회지 10(1), pp.44-63.

박종수·이금숙, 2008, “서울대도시권 지하철망의 구조적 특성분석,” 한국경제지리학회지 11(3), pp.459-475.

박종수·이금숙, 2011, “Pickup Point 최적입지선정을 위한 Greedy Heuristic Algorithm 개발 및 적용: 서울 대도시권 지하철 시스템을 대상으로,” 한국경제지리학회지 14(2), pp.116-128, 2011.

이금숙·박종수·김호성·조창현, 2010, “수도권 광역도시 철도 하루 시간대별 이용 빈도에 의해 구분된 역 집단과 통행자의 통행 연쇄 패턴 간 관계,” 대한지리학회지 45(5), pp.592-608.

이금숙·박종수, 2006, “서울시 대중교통 이용자의 통행 패턴 분석,” 한국경제지리학회지 9(3), pp.379-395.

서울 메트로 노선도 서비스, 2012, <http://www.seoul-metro.co.kr/station/linemap.action>(최종열람일: 2012년 7월 10일)

Assunção, R.M., Neves, M.C., Camara, G. and Freitas, C.D.C., 2006, “Efficient Regionalization Techniques for Socio-Economic Geographical Units Using Minimum Spanning Trees,” *International Journal of Geographical Information Science* 20(7), pp.797-811.

Chen, C., Chen, J. and Barry, J., 2009. Diurnal pattern of transit ridership: a case study of the NewYorkCity-subwaysystem. *Journal of Transport Geography* 17, pp.176-186.

Delling, D., Goldberg, A.V., Razenshteyn, I. and Werneck, R.F., 2010, “Graph Partitioning with Natural Cuts,” *Microsoft Technical Report*, MSR-TR-2010-164.

Goh, S., Lee, K., Park, J. S., and Choi, M. Y., 2012, “Modification of the gravity model and application to the metropolitan Seoul subway system,” DOI: 10.1103/PhysRevE.86.026102.

Guo, D., 2009, “Flow Mapping and Multivariate Visualization of Large Spatial Interaction Data,” *IEEE Transactions on Visualization and Computer Graphics* 15(6), pp.1041-1048.

Karypis, G., 2011, “METIS: A Software Package for Partitioning Unstructured Graph, Partitioning Meshes, and Computing Fill-Reducing Orderings of Sparse Matrices,” Version 5.0, <http://www.cs.umn.edu/~metis>, 34 pages.

Karypis, G. and Kumar, V., 1988, “A Fast and High Quality Multilevel Scheme for Partitioning Irregular Graphs,” *SIAM Journal on Scientific Computing* Volume: 20, Issue: 1, pp.359-392.

Lee, Keumsook , Song, Yena, Park, Jong Soo, Anderson, William P., 2012, “Relationship between Diurnal Patterns of Transit Ridership and Land Use in the Metropolitan Seoul Area,” *Journal of the Economic Geographical Society of Korea* 15(1), pp.26-41.

Lee, K., Goh, S., Park, J.S., Jung, W.-S., and Choi, M.Y.,

2011, "Master equation approach to the intra-urban passenger flow and application to the Metropolitan Seoul Subway System," *Journal of Physics A: Mathematical and Theoretical*, 44(2011) 11507(14pp).

Roth, C., Jang, S. M., Batty, M., and Barthelemy, M., 2011, Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows, *PLoS ONE*, 6(1), e15923.

Tan, P.-N., Steinbach, M. and Kumar, V., 2006, *Introduction to Data Mining*, Boston: Addison-Wesley,

pp.548-555.

교신: 이금숙, 136-742, 서울특별시 성북구 동선동 3가 249-1, 성신여자대학교 지리학과, 전화: 02-920-7138, 이메일: kslee@sungshin.ac.kr

Correspondence: Keumsook Lee, 249-1 Dongseon-dong 3-ga, Seongbuk-gu, Seoul 136-742, Korea, Tel: +82-2-920-7138, e-mail: kslee@sungshin.ac.kr

최초투고일 2012년 8월 11일

최종접수일 2012년 8월 27일

Journal of the Economic Geographical Society of Korea
Vol.15, No.3, 2012(343-357)

Classification of the Seoul Metropolitan Subway Stations using Graph Partitioning

Jong Soo Park* · Keumsook Lee**

Abstract : The Seoul metropolitan subway system can be represented by a graph which consists of nodes and edges. In this paper, we study classification of subway stations and trip behaviour of subway passengers through partitioning the graph of the subway system into roughly equal groups. A weight of each edge of the graph is set to the number of passengers who pass the edge, where the number of passengers is extracted from the transportation card transaction database. Since the graph partitioning problem is NP-complete, we propose a heuristic algorithm to partition the subway graph. The heuristic algorithm uses one of two alternative objective functions, one of which is to minimize the sum of weights of edges connecting nodes in different groups and the other is to maximize the ratio of passengers who get on the subway train at one subway station and get off at another subway station in the same group to the total subway passengers. In the experimental results, we illustrate the subway stations and edges in each group by color on a map and analyze the trip behaviour of subway passengers by the group origin-destination matrix.

Key Words : Seoul metropolitan subway system, transportation card transaction database, trip behaviour of subway passengers, graph partition, origin-destination matrix

* professor, School of Information Technology, Sungshin Women's University

** professor, Department of Geography, Sungshin Women's University