

# Graph Centrality를 통한 최적의 상권 입지 요건 분석: 스타벅스를 중심으로

송근영(2017103997)

## 요약

Graph Centrality와 다양한 알고리즘으로 빅데이터를 분석해 최적의 상권 입지 요건을 파악할 수 있는 프로그램을 설계한다. 그 이후, 프로그램을 통해 도출된 데이터를 분석해 시사하는 바를 파악한다.

## 1. 서론

### 1.1 연구배경

창업을 해 입점을 하기에 앞서, 해당 상권을 분석하는 것은 매우 중요하다. 무턱대고 차렸다가는 쉽게 망할 수 있다. 따라서, 입점하고자 하는 점포 주변 상권을 과학적이고 체계적인 기법으로 파악하여 최적의 장소인지를 분석할 필요가 있다. 예를 들어 주변에 베이커리 가게가 5곳 있는데, 이곳에 똑같이 베이커리 가게를 창업한다면 망할 가능성은 매우 높을 것이다.

하지만, 아직 이러한 상권 분석에 대한 과학적인 연구 사례는 많지 않다. 아직도 많은 점포가 입지의 결함으로 많은 폐업을 반복하고 있다. 따라서, 기회비용을 따지며 좋은 위치에 자리 잡는 것은 매우 중요하다. 이때, 이미 입점에 성공해 안정화된 빅데이터를 종합해 분석한다면 지리적 요소에 따른 사업 흥망 여부를 예측할 수 있을 것이며, 이를 시스템화 한다면 빠르고 간단하게 좋은 요건을 분석하는 것이 가능하다. 따라서 본 연구에선 Graph Centrality 기법을 이용해 빅데이터를 분석해 선정한 지역에 해당 업종이 적절한지 알 수 있는 시스템 기반을 설계하고자 한다.

### 1.2 연구목표

본 연구에선 Graph Centrality 기법을 이용해 빅데이터를 분석해서 선정한 지역에 해당 업종이 적절한지 알 수 있는 시스템 기반을 설계하고, 해당 결과가 시사하는 바를 분석하고자 한다. 예를 들어, 카페 업종을 입력하고 나온 데이터를 분석한다면 고등학교 주변보다 대학교 주변에서 해당 업종이 흥할 가능성이 높다는 것을 파악할 수 있다. 그 이후, 이러한 수치를 통계 내, 해당 업종이 그 지점에서 얼마나 뛰어난 지리적 입점 점수를 받을 수 있는지 계산하고 추천한다.

### 1.3 표본데이터 선정

본 연구에서 대표 표본 데이터로 스타벅스를 선정했다. 일단 스타벅스는 ‘스타벅스 효과’라는 용어를 지니고 있을 만큼 좋은 상권에 입점한다. 스타벅스는 모든 매장이 직영점으로 운영돼, 일반인은 창업할 수 없고, 건물주만 스타벅스 입점 의뢰가 가능하다. 의뢰를 받은 경우, 스타벅스 본점이 ‘주변 스타벅스까지와의 거리’, ‘타 경쟁업체’, ‘역과의 거리’, ‘정류장과의 거리’ 등 여러 지리적 요소를 수집해 승인 여부를 결정한다. 따라서, 해당 연구 표본으로 가장 신뢰 있는 데이터라 판단했고, 대표 표본으로 선정하였다.

## 2. 관련연구

### 2.1 Graph Centrality

그래프에서 중심성이란 그래프 혹은 사회 연결망에서 꼭짓점 혹은 노드의 상대적 중요성을 나타내는 척도이다. 이 중심성은 지수로 계산되는데 기준에 따라서 다양하다. 노드들 간의 최단 경로를 가지고 계산하는 매개 중심성(Betweenness centrality), 그래프의 노드와 다른 모든 노드들 사이의 최단 경로 길이의 합의 역수를 이용해서 구하는 근접 중심성(Closeness Centrality), 한 노드에 연결된 모든 엣지의 개수로 중심성을 평가하는 연결 중심성(Degree Centrality) 등 이 밖에도 다른 계산 방법들이 많이 존재한다.

### 2.2 참고 연구

관련 연구로는 ‘스타벅스 입지의 공간적 효과에 관한 연구, 2016년 개점된 서울 시내 점포를 대상으로 논문’이 있다. 해당 연구에선 스타벅스가 들어서서 주변 상권이 좋아지는 것은 증명되지 않은 사실이며, 기존에 이미 상권이 좋은 곳에 스타벅스가 입점한다는 사실을 도출했다. 이를 근거 삼아 본 연구의 전제인 ‘스타벅스는 좋은 상권에 입점하므로 대표 표본 데이터로 선정한다.’를 해당 논문으로 증명하며, 이를 기반으로 연구를 진행한다.

## 3. 프로젝트

### 3.1 프로젝트 내용

웹 크롤링을 통해 대량의 스타벅스 위치 데이터를 불러온다. 이후, 대학교, 역, 정류장 등의 위치포인트나 유동인구, 소득수준, 주변 경쟁업체 수 등 수집해 데이터화한다. 그 다음, 스타벅스 매장과 다른 위치 데이터를 노드로 선정하고, 각 요소별 수치를 이용해 엣지에 가중치를 적절하게 부여해 네트워크를 형성한다. 이때, 거리가 가깝거나 해당 수치가 높으면 엣지가 굵어지며, 이에 따라 각 노드의 크기가 커지게 된다. 이렇게 제작된 그래프를 Graph Centrality를 이용해 시각화해 분석하고, 각 지역의 구역을 점수화해 선점한 지역이 지리적 요소로 얼마

나 적합한지 손쉽게 파악할 수 있도록 한다.

## 3.2 요구사항

### 3.2.1 정제된 데이터

위와 같은 프로젝트를 위해 데이터 수집 및 정제가 필요하다. 본 프로젝트는 스타벅스에서 제공하는 위치 데이터와 구글 맵 API를 이용하여, 대학교, 역, 정류장 등의 위치포인트 위도와 경도를 불러온다. 또한, 유동인구, 소득수준, 주변 경쟁업체 등은 웹 크롤링을 통하여 수집한다. 이후 시각화 도구 ‘Gephi’와 호환될 수 있도록 데이터를 교정 및 통합해 정제한다.

### 3.2.2 Graph Centrality

본 프로젝트를 수행하기 위한 핵심 알고리즘이다. 노드와 엣지로 표현된 그래프에서 노드들의 상대적 중요성을 계산하는 기법이다. 계산하는 방법에 따라 크게 매개 중심성(Betweenness Centrality), 연결 중심성(Degree Centrality), 근접 중심성(Closeness centrality), 고유벡터 중심성(Eigenvector Centrality)로 분류된다. 현재 프로젝트는 근접 중심성을 주로 사용하되, 상황에 맞게 다른 중심성 기법을 사용해 각자 다르게 출력된 결과를 대조해 분석한다. 근접 중심성이란 각 노드 간의 거리를 바탕으로 중심성을 측정하는 방식이다. 근접 중심성은 간접적으로 연결되어 있는 모든 노드간의 거리를 바탕으로 중심성을 측정한다. 직접 연결되어 있는 사람들을 통해 접근할 수 있는 간접적인 노드 간의 관계까지 파악하며, 연결정도 중심성보다 넓은 범위에서 중심성을 측정하게 된다. 간접적인 중심성이 크다면, 핵심 노드와의 관계가 많다는 뜻으로 다수의 핵심 노드들과의 관계를 통해 해당 네트워크 내의 핵심 노드일 가능성이 높다.

### 3.2.3 Gephi

대규모 그래프를 시각화하고 분석하기 위한 오픈 소프트웨어이다. 3D 렌더링 엔진을 사용해 실시간으로 그래프를 표시하고 탐색 속도를 향상해준다.

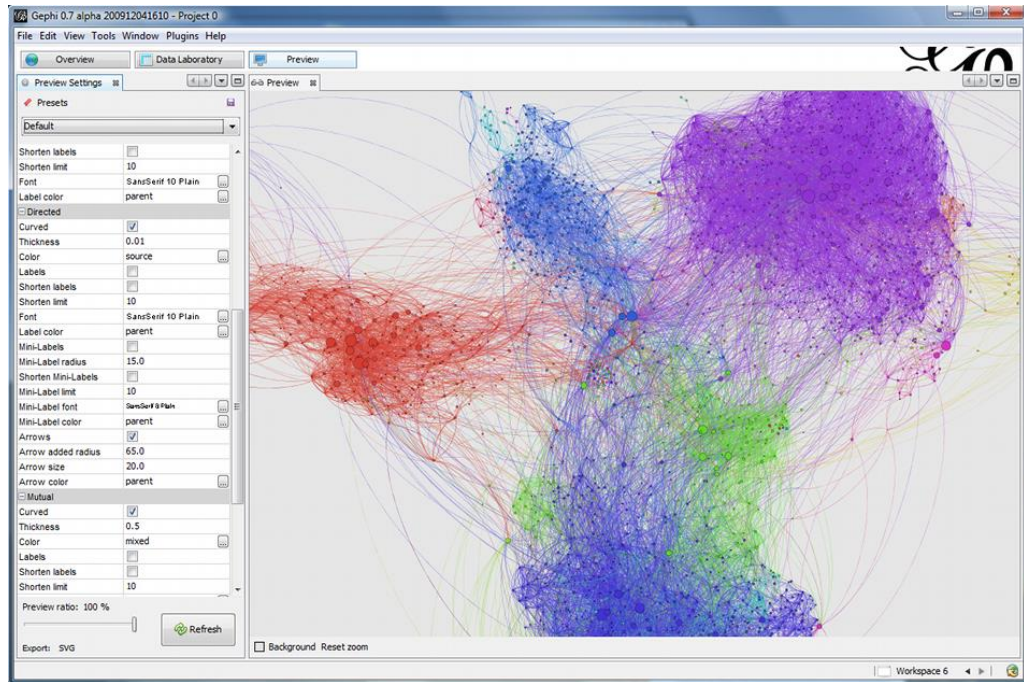


그림 2 [Gephi 화면 예시]

### 3.3 개발환경

운영체제: 윈도우 11

개발언어: Python

개발도구: Gephi, Igraph, NetworkX

## 4. 향후 일정 및 역할 분담

진행 일정	내용
~ 10/5	프로젝트 데이터 수집
10/5 ~ 10/25	데이터 정제 및 알고리즘 작성
10/25 ~ 11/1	그래프 중심성 측정
11/1 ~ 11/10	중심성 기반 분석
11/10 ~ 11/18	새로운 데이터 적용 및 반복
11/18 ~ 11/25	최종보고서 작성 및 제출
11/25 ~ 12/5	발표 준비 및 발표

## 5. 결론 및 기대효과

앞선 빅데이터를 분석해 유사 업종은 어떤 지리적 요소에서 성공했는지 손쉽게 파악할 수 있다. 이를 토대로 입점하고자 하는 지역을 해당 프로그램으로 분석해 지리적 요소로 인한 흥망 여부를 예측할 수 있다. 또한, 도출된 데이터를 분석함으로써 해당 업종이 지닌 특징과 시사하는 바를 파악할 수 있다.

## 6. 참고문헌

김흥순, 김영덕, 원윤재 and 신은하. (2020). 스타벅스 입지의 공간적 효과에 관한 연구 - 2016년 개점된 서울시내 점포를 대상으로 -. 국토지리학회지, 54(1), 77-89.