캡스톤 11회차 연구노트

네트워크 분석을 하는 이유는?

각 지역에서 언급되는 키워드의 연관성을 파악하여 브랜딩 전량이나 정책 제안 등에 활용하기 위해

본 프로젝트의 네트워크 분석에서 노드, 간선, 가중치의 의미는?

노드: 웹 텍스트에서 추출된 키워드(명사)

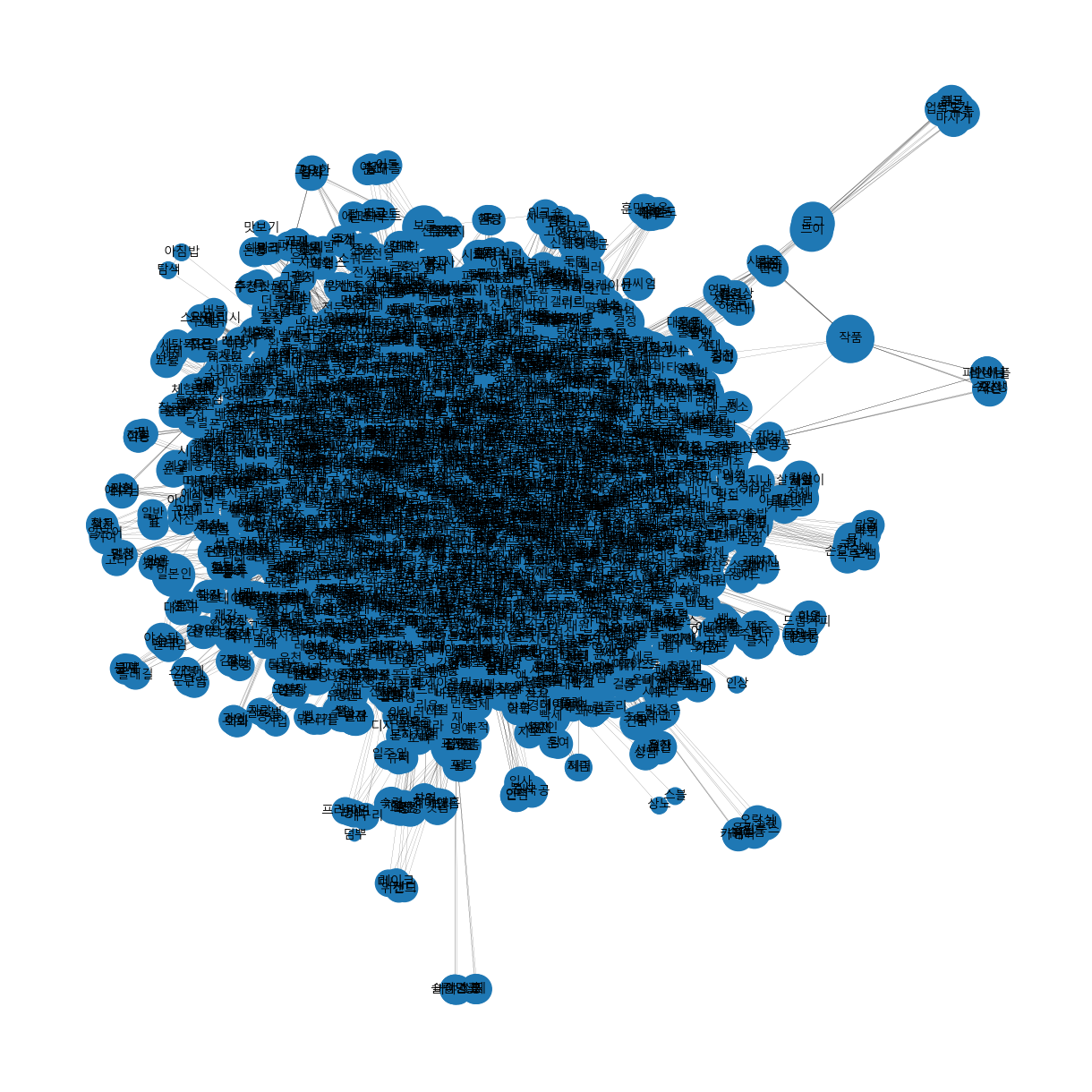
간선: 서로 다른 두 노드의 연결(한 문서에서의 공동 출현 여부)

예) 1번 문서에서 N번 문서 중 ‘맛집’과 ‘서문시장’이 공동으로 포함된 문서가 있다면 두 키워드 사이의 간선이 있음

가중치: 각 지역마다의 총 문서에 등장한 공동 출현 횟수의 총합

예) 대구 관련 텍스트 모음집에서 ‘맛집’과 ‘서문시장’이 공동으로 포함된 문서가 20개이면 해당 노드 간 간선의 가중치는 20

문서 당 키워드를 정하는 방법은(개수, 불용어 처리 등)?



문서 수 자체가 많기 때문에 문서 당 상위 최대 10개만 수집해도 너무 밀집되어 있음

* 동시 출현 빈도가 높은 쌍 100개를 대상으로 네트워크 구축

원, 스크린샷이(가) 표시된 사진

자동 생성된 설명

각 키워드의 중심성을 엑셀 파일로 정리

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **words** | **degree** | **betweenness** | **closeness** | **eigenvector** |
| 맛집 | 0.38 | 0.34 | 0.42 | 0.46 |
| 호텔 | 0.25 | 0.19 | 0.33 | 0.36 |
| 추천 | 0.15 | 0.12 | 0.39 | 0.40 |
| 동성로 | 0.12 | 0.07 | 0.37 | 0.34 |
| 막창 | 0.05 | 0.04 | 0.32 | 0.13 |
| 숙소 | 0.11 | 0.03 | 0.30 | 0.23 |
| 수성 | 0.03 | 0.03 | 0.33 | 0.06 |
| 동대구역 | 0.03 | 0.03 | 0.33 | 0.09 |
| 서문시장 | 0.03 | 0.02 | 0.27 | 0.05 |
| 식당 | 0.03 | 0.02 | 0.27 | 0.07 |
| 거리 | 0.03 | 0.02 | 0.27 | 0.03 |
| 카페 | 0.08 | 0.01 | 0.32 | 0.23 |
| 코스 | 0.08 | 0.01 | 0.32 | 0.22 |
| 예약 | 0.08 | 0.00 | 0.23 | 0.15 |
| 포토 | 0.04 | 0.00 | 0.04 | 0.00 |
| 아고다 | 0.08 | 0.00 | 0.23 | 0.21 |
| 스테이 | 0.05 | 0.00 | 0.23 | 0.06 |
| 라이온즈 | 0.03 | 0.00 | 0.03 | 0.00 |
| 데이트 | 0.07 | 0.00 | 0.32 | 0.14 |
| 할인 | 0.07 | 0.00 | 0.23 | 0.20 |
| 교동 | 0.03 | 0.00 | 0.30 | 0.07 |
| 커피 | 0.03 | 0.00 | 0.27 | 0.05 |
| 필수 | 0.03 | 0.00 | 0.27 | 0.05 |
| 고기 | 0.03 | 0.00 | 0.27 | 0.04 |
| 뭉티기 | 0.03 | 0.00 | 0.27 | 0.04 |
| 당일치기 | 0.03 | 0.00 | 0.27 | 0.04 |
| 중구 | 0.01 | 0.00 | 0.26 | 0.03 |
| 갈비 | 0.01 | 0.00 | 0.26 | 0.04 |
| 시장 | 0.01 | 0.00 | 0.26 | 0.02 |
| 칼국수 | 0.01 | 0.00 | 0.26 | 0.02 |
| 웨이 | 0.01 | 0.00 | 0.26 | 0.03 |
| 부산 | 0.01 | 0.00 | 0.26 | 0.02 |
| 방문 | 0.01 | 0.00 | 0.26 | 0.04 |
| 후기 | 0.01 | 0.00 | 0.26 | 0.05 |
| 떡볶이 | 0.01 | 0.00 | 0.26 | 0.02 |
| 진짜 | 0.01 | 0.00 | 0.26 | 0.04 |
| 수성구 | 0.01 | 0.00 | 0.26 | 0.04 |
| 디저트 | 0.03 | 0.00 | 0.26 | 0.04 |
| 명소 | 0.01 | 0.00 | 0.25 | 0.02 |
| 여행지 | 0.01 | 0.00 | 0.25 | 0.03 |
| 소품 | 0.01 | 0.00 | 0.25 | 0.02 |
| 코드 | 0.05 | 0.00 | 0.23 | 0.13 |
| 가성 | 0.03 | 0.00 | 0.23 | 0.05 |
| 위치 | 0.03 | 0.00 | 0.23 | 0.07 |
| 노블 | 0.03 | 0.00 | 0.23 | 0.03 |
| 인터 | 0.01 | 0.00 | 0.23 | 0.04 |
| 호캉스 | 0.01 | 0.00 | 0.23 | 0.02 |
| 메리어트 | 0.01 | 0.00 | 0.23 | 0.03 |
| 숙박 | 0.01 | 0.00 | 0.23 | 0.03 |
| 시설 | 0.01 | 0.00 | 0.23 | 0.03 |
| 걸리버 | 0.01 | 0.00 | 0.22 | 0.01 |
| 구공탄 | 0.01 | 0.00 | 0.22 | 0.01 |
| 신라 | 0.01 | 0.00 | 0.20 | 0.00 |
| 김광석 | 0.01 | 0.00 | 0.20 | 0.00 |
| 먹거리 | 0.01 | 0.00 | 0.20 | 0.00 |
| 스페셜 | 0.03 | 0.00 | 0.03 | 0.00 |
| 덤프 | 0.03 | 0.00 | 0.03 | 0.00 |
| 일상 | 0.01 | 0.00 | 0.02 | 0.00 |
| 삼성 | 0.01 | 0.00 | 0.02 | 0.00 |
| 파크 | 0.01 | 0.00 | 0.02 | 0.00 |
| 추석 | 0.01 | 0.00 | 0.01 | 0.00 |
| 근대 | 0.01 | 0.00 | 0.01 | 0.00 |
| 백화점 | 0.01 | 0.00 | 0.01 | 0.00 |
| 간송 | 0.01 | 0.00 | 0.01 | 0.00 |
| 공원 | 0.01 | 0.00 | 0.01 | 0.00 |
| 미술관 | 0.01 | 0.00 | 0.01 | 0.00 |
| 언덕 | 0.01 | 0.00 | 0.01 | 0.00 |
| 경상감영 | 0.01 | 0.00 | 0.01 | 0.00 |
| 신세계 | 0.01 | 0.00 | 0.01 | 0.00 |
| 연휴 | 0.01 | 0.00 | 0.01 | 0.00 |
| 중화 | 0.01 | 0.00 | 0.01 | 0.00 |
| 청라 | 0.01 | 0.00 | 0.01 | 0.00 |
| 비빔밥 | 0.01 | 0.00 | 0.01 | 0.00 |
| 골목 | 0.01 | 0.00 | 0.01 | 0.00 |

어떤 중심성을 활용할지(조합하는 경우도 염두)에 따라 관광 활성화 정책을 달리 모색할 수 있음

예) 연결 중심성 -> 허브 지역을 중심으로 쉽게 방문할 수 있는 주변 관광지에 집중하기

매개 중심성 -> 다른 지역과 연결되는 곳을 중심으로 비교적 넓은 지역적 범위를 고려하기