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)

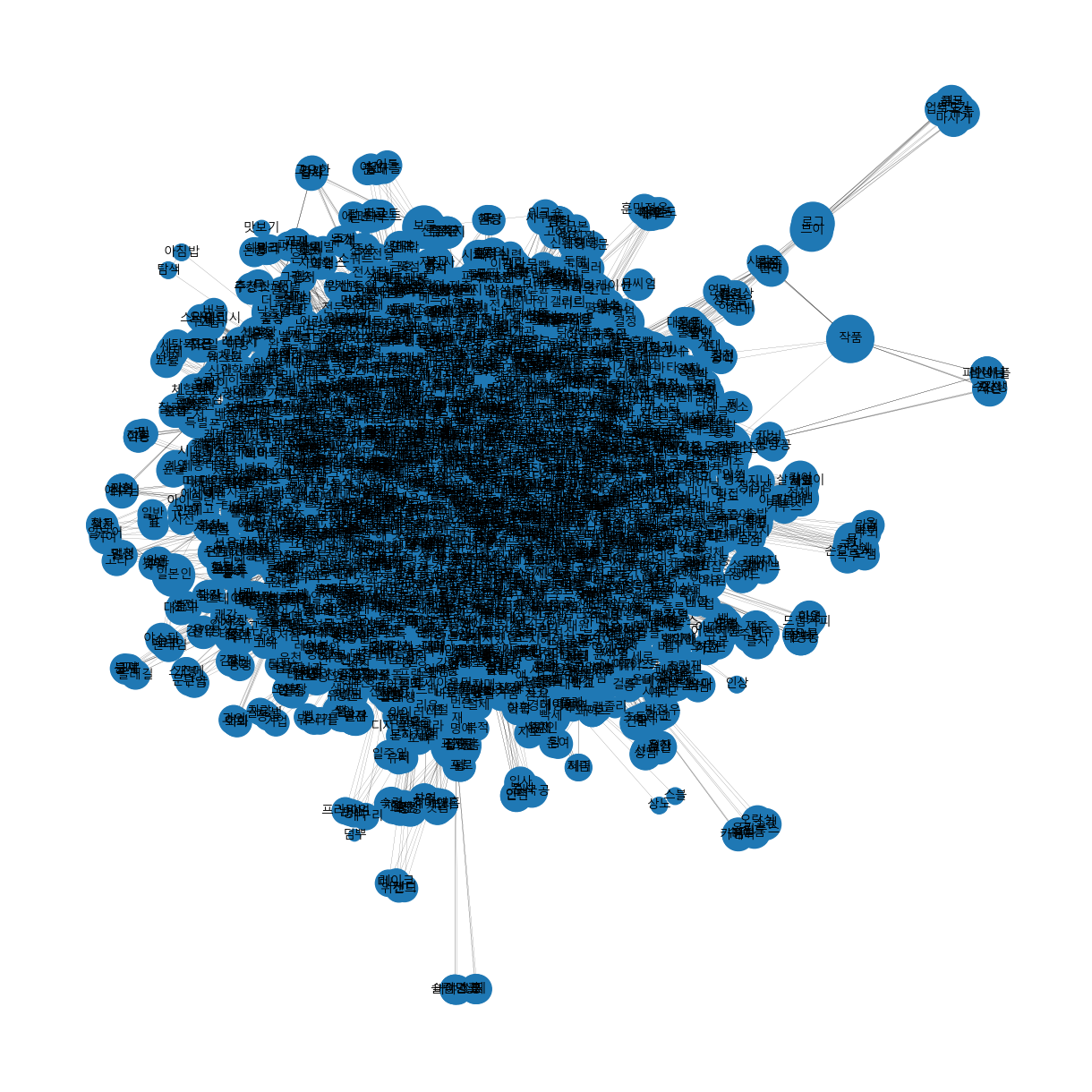
간선: 서로 다른 두 노드의 연결(한 문서에서의 공동 출현 여부)

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)?

그림 1

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

* 동시 출현 빈도가 높은 쌍 100개를 대상으로 네트워크 구축

**대구**

원, 스크린샷이(가) 표시된 사진

자동 생성된 설명그림 2

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 |

(후략)

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

근접 중심성(특정 노드에서 다른 노드까지의 최소 경로의 평균의 역수) -> 값이 클수록 전체 네트워크에서 중심적인 포지션 -> 해당 지역의 대표적인 키워드에 집중하기

한편 '맛집', '호텔', '숙소', '추천', '후기', '코스'와 같이 여행과 관련된 일반적인 키워드를 분석에서 제외한 경우 다음과 같은 결과가 도출되었다.

원, 그림, 대칭이(가) 표시된 사진

자동 생성된 설명그림 3

[그림 2]와는 달리 좀 더 분산적인 모습을 띠고 있다.

분석 방법이 어느정도 정해졌기 때문에 대전, 울산 여행에 관한 키워드 네트워크 생성은 어렵지 않았다.

**대전**

원이(가) 표시된 사진

자동 생성된 설명그림 4

위 그래프는 대전 여행에 관한 키워드의 네트워크이다. 한눈에 봐도 성심당의 위엄이 드러난다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **words** | **degree** | **betweenness** | **closeness** | **eigenvector** |
| 성심당 | 0.5 | 0.564941 | 0.518006 | 0.609245 |
| 호텔 | 0.185714 | 0.192754 | 0.320829 | 0.030782 |
| 맛집 | 0.157143 | 0.09648 | 0.394671 | 0.347942 |
| 추천 | 0.1 | 0.218703 | 0.41098 | 0.310485 |
| 코스 | 0.085714 | 0.017598 | 0.373899 | 0.2804 |
| 예약 | 0.071429 | 0.000828 | 0.237936 | 0.008727 |
| 도시 | 0.057143 | 0.047619 | 0.333749 | 0.098277 |
| 숙소 | 0.057143 | 0.019048 | 0.301385 | 0.035172 |
| 카페 | 0.057143 | 0.003589 | 0.36836 | 0.216685 |
| 시루 | 0.057143 | 0.000621 | 0.329328 | 0.199045 |
| 아고다 | 0.057143 | 0.000138 | 0.236803 | 0.007503 |
| 할인 | 0.057143 | 0.000138 | 0.236803 | 0.005358 |
| 포토 | 0.042857 | 0.047619 | 0.246181 | 0.002813 |
| 칼국수 | 0.042857 | 0.024017 | 0.33829 | 0.139285 |
| 롯데 | 0.042857 | 0.011939 | 0.287448 | 0.010936 |
| 당일치기 | 0.042857 | 0.001035 | 0.333749 | 0.145105 |

(후략)

중심성 분석에서 정규화 된 근접 중심성이 0.5인 점을 보니 그 정도를 더 확실하게 느낄 수 있었다. 한편 카페 역시 높은 편에 속한다.

내륙 도시 특성상 자연 경관보다는 도시 여행을 테마로 관광 상품(홍보)을 모색하는 것이 나을 것으로 보인다(시티투어, 상점가, 감성카페 등).

**울산**

**원이(가) 표시된 사진

자동 생성된 설명**그림 5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **words** | **degree** | **betweenness** | **closeness** | **eigenvector** |
| 추천 | 0.206349 | 0.402884 | 0.457294 | 0.462805 |
| 맛집 | 0.31746 | 0.334443 | 0.440513 | 0.520865 |
| 코스 | 0.126984 | 0.21659 | 0.390373 | 0.274847 |
| 대왕암 | 0.095238 | 0.018348 | 0.366533 | 0.282448 |
| 공원 | 0.095238 | 0.018348 | 0.366533 | 0.265406 |
| 데이트 | 0.063492 | 0.006827 | 0.366533 | 0.193314 |
| 카페 | 0.111111 | 0.086278 | 0.350481 | 0.221747 |
| 태화강 | 0.079365 | 0.106503 | 0.350481 | 0.113231 |
| 호텔 | 0.190476 | 0.236559 | 0.347941 | 0.172337 |
| 바다 | 0.047619 | 0.004608 | 0.333444 | 0.123793 |
| 숙소 | 0.047619 | 0.011265 | 0.326639 | 0.118897 |
| 추석 | 0.063492 | 0.040963 | 0.311791 | 0.052995 |
| 선물 | 0.047619 | 0.013313 | 0.30978 | 0.055752 |
| 간절곶 | 0.031746 | 0 | 0.301987 | 0.096733 |

(후략)

울산은 해안 경관의 중요도가 높은 것으로 나타났으며(대왕암, 간절곶) 다른 두 지역에 비해 ‘데이트’의 근접 중심성이 높았다.