

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

Roaming Call Data를 통한 새로운 관광 정책/마케팅 제안 기법 - AR, SNA를 중심으로

---

제 2회 관광 빅 데이터 분석 대회 (기획 부문)



2014.04.30

## 목차

---

1. 프로젝트 개괄
2. 현상 분석
  - 1) 기존 분석 전략의 한계
  - 2) 외래관광객의 불만요소 분석
3. 제공 데이터 파악
4. 프로젝트의 목적·목표
  - 1) 연관 분석 (Association Rules)
  - 2) 사회 연결망 분석 (SNA)
  - 3) 기존 분석과의 비교 및 분석 적합성 평가
  - 4) 요약
5. 프로젝트 일정
6. 참고문헌

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

- 1. 프로젝트 개괄
- 2. 현상 분석
- 3. 제공 데이터 파악
- 4. 프로젝트의 목적·목표
- 5. 프로젝트 일정

## 1. 프로젝트 개괄

▶ 전체 프로젝트는 다음과 같이 진행될 계획이며, 기획부문에서는 과제 목적·목표 단계까지 진행하였음

### 현상 분석

- 관광 사업의 정책 / 마케팅적 현황 조사

### 데이터 파악

- 제공된 Raw 데이터 기본적인 의미 파악

### 과제 목적·목표

- 분석 방법의 결정과 예상 분석 결과의 의미 해석
- 예상 결과의 의미를 사회 문제와 연결
- 분석 적합성 평가

### 데이터 분석

- 연관 분석(AR)
  - \* SAS 9.2 사용
  - \* SAS E-Miner 사용
- 사회 연결망 분석(SNA)
  - \* UCINET6 사용

### 결과 해석

- 분석 결과 해석
- 구체적인 정책 전략 및 마케팅 전략 제시

### 기획 부문에서의 수행 과제

- 현상 분석 단계에서는 '정책/마케팅 측면에서의 현황'을 조사
- 데이터 파악 단계에서는 '제공된 Raw 데이터의 기본적인 의미'를 파악
- 과제 목적·목표 단계에서 'Raw 데이터를 분석해서 나올 결과의 의미'를 예측하고 적용 가능한 정책/마케팅 분야를 설정. 이를 바탕으로 기존 분석과의 비교 및 적합성 평가
- 데이터 분석 단계 / 결과 해석 단계는 분석 부문에서 수행

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

- 1. 프로젝트 개괄
- 2. 현상 분석
- 3. 제공 데이터 파악
- 4. 프로젝트의 목적·목표
- 5. 프로젝트 일정

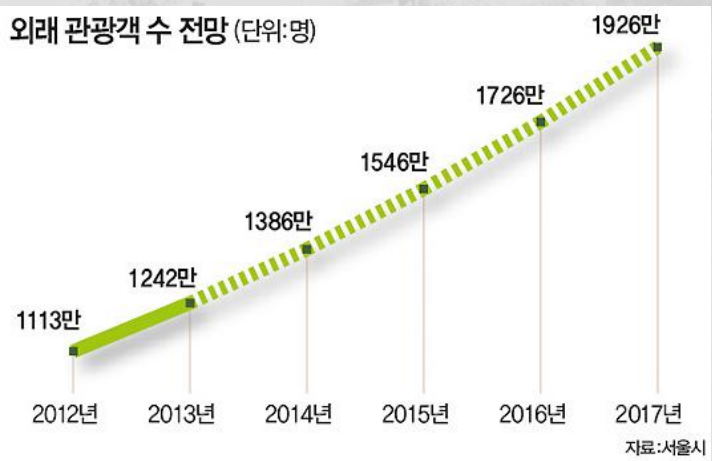
## 2. 현상 분석 1) 기존 분석 전략의 한계

### ▶ 빅 데이터를 기반으로 한 관광 정책/마케팅의 미비

- 대한민국은 2013년 한해 동안, 1200만 명이 넘는 외래 관광객이 방문하였고, 다양한 해외관광 마케팅을 시행하고 있다.([그림 1] 참고)
- 정책/마케팅의 기반이 되는 자료 조사는 전체 관광객의 0.1% 정도인 약 12,000명(매월 약 1,000명)에게만 설문조사를 진행하는 것으로 그친다.<sup>1)</sup>

- 시장별 유치전략 수립
- 방한관광상품 개발, 판촉
- 초청지원(여행업자, 언론인, 유관인사 등)
- 소비자 대상 홍보이벤트 개최, 전시박람회 참가 한국 홍보
- 신규, 잠재시장 개척
- MICE 등 대량 인센티브관광객 유치
- 해외 광고(TV, 인쇄매체, 온라인 등)

[그림 1] 한국관광공사-해외관광마케팅



[그림 2] 외래관광객 수 전망

- 반면, 신한카드에서 관광객 행태를 분석하여 유용한 정보를 제공하고 있다.<sup>2)</sup>
- 갈수록 늘어나는 외래관광객의 의견을 모두 설문조사로 파악하기론 한계가 있다.
- 기존 설문조사를 통한 방식에서 탈피한 실증적 빅 데이터에 대한 이해, 분석 기법의 필요성 증가되고 있다.<sup>3)</sup>

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

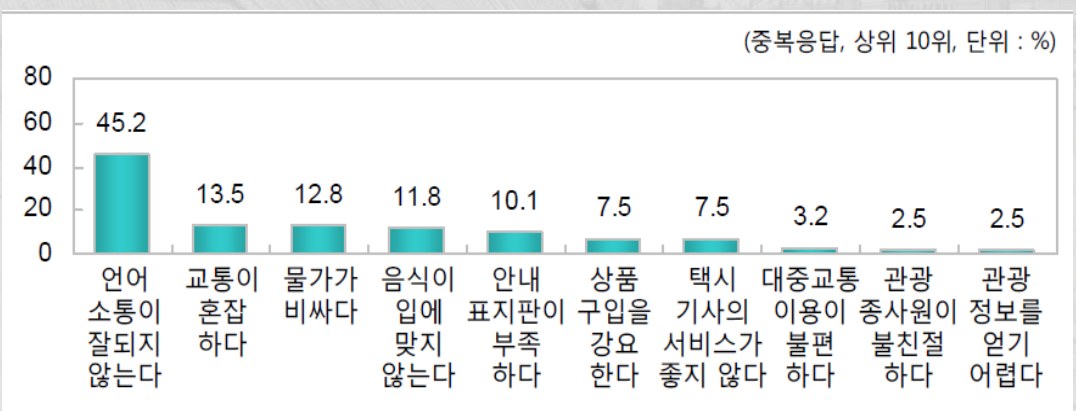
1. 프로젝트 개괄
2. 현상 분석
3. 제공 데이터 파악
4. 프로젝트의 목적·목표
5. 프로젝트 일정

## 2. 현상 분석 2) 외래관광객의 불만요소 분석

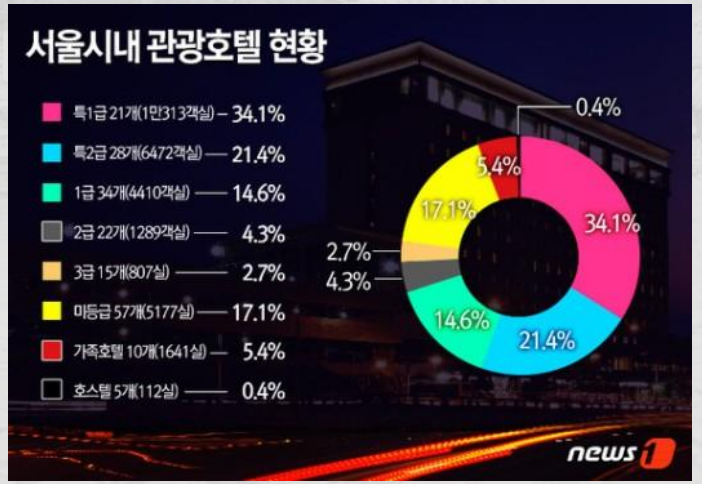
### ▶ 중저가 숙박시설의 부족 및 언어, 교통 문제

- 외래방문객의 가장 큰 불만요소는 언어, 교통, 물가, 음식 순이었다.([그림 3] 참고) <sup>1)</sup>
- 전반적인 숙박 수요 전망은 앞으로도 계속 증가하지만 공급은 이를 따라가지 못하는 현실이다.<sup>4)</sup>
- 서울의 경우, 특급호텔이 전체 객실의 절반을 넘는 55%이지만, 외국인 관광객들이 선호하는 2, 3급 호텔은 7%수준에 그치고 있다.<sup>5)</sup>
- '비싼 물가'에서 숙박료가 많은 부분을 차지하고 이에 대한 해결책이 필요한 시점이다.
- 중저가 숙박시설을 활성화할 수 있는 보다 현실적인 방법이 필요하다.

◇“외국인 관광객 절반 ‘숙박료 비싸서’ 한국 방문 포기”



[그림 3] 한국문화관광연구원. 2013년도 외래관광객의 한국 여행시 불편사항



[그림 4] 파이낸셜뉴스. 중저가 관광호텔 필요한데 특급만 '줄줄' (2014-04-03)

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

1. 프로젝트 개괄
2. 현상 분석
3. 제공 데이터 파악
4. 프로젝트의 목적·목표
5. 프로젝트 일정

## 3. 제공 데이터 파악

### ▶ 제공된 데이터 셋(기간 내의 국내 로밍 통화/SMS 발생 기록)의 의미

Column	설명	척도 유형	중복 제외 데이터 수	추가 설명
ID	암호화된 국내 로밍 고객 구분 코드	Nominal	273,553 (분석할 관광객 수)	1인 평균 6.5(=1,774,618/273,553)회 통화/SMS를 사용
DATE	Call 발생 날짜 (연월일)	Interval	생략	국내 기념일(크리스마스, 설날 등), 해외 기념일(중국 연휴 등)등을 추출하여 분석 가능 (2013.12.01~2014.02.28)
TIME	Call 발생 시간 (시분)	Interval	생략	
COUNTRY_CODE	로밍 국가 코드 (N01~N10 : 10개국)	Nominal	10	국가별로 비교 분석 가능
ZIP	Call 발생 기지국 위치(우편번호)	Nominal	5,746	<b>Call 발생 지역을 여행 방문 장소로 생각할 수 있음</b>  우편 번호를 이용하여 지역별(e.g. XX시 XX구 XX동)로 분석 가능
LATITUDE	Call 발생 기지국 위치(위도)	Interval	9,310 (위도, 경도) 측정	
LONGITUDE	Call 발생 기지국 위치(경도)	Interval		

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

1. 프로젝트 개괄 2. 현상 분석 3. 제공 데이터 파악 4. 프로젝트의 목적·목표 5. 프로젝트 일정

## 4. 프로젝트의 목적 · 목표 1) 연관 분석(Association Rules) – 설명 및 예시

▶ 연관 분석을 통해 외래 관광객들의 중요 관광지 코스를 파악 할 수 있음

### •연관 분석(Association Rules) 이란?

대표적인 발견 위주의 데이터 마이닝으로 고객의 구매 상품 간의 관계를 파악하는 데 많이 활용되는 기술이다. 결과가 'A → B'라면, '상품 A를 구매하면 동시에 상품 B도 구매한다'라고 해석할 수 있다.<sup>6)</sup>

•ID 변수 : ID(암호화된 국내 로밍 고객 구분 코드)

•ITEM 변수 : ZIP (Call 발생 기지국 우편번호)

•가상 결과 : 618807(부산광역시 강서구 대저2동) → 690801(제주특별시 제주시 건입동) (confidence : 90%)

•해석 : 부산광역시 강서구 대저2동을 간 관광객의 90%가 제주특별시 제주시 건입동을 방문한다.

•가능한 정책 : 대저2동 관광지에 제주도 관광지 안내도 혹은 안내 표지판을 설치한다.

•가능한 마케팅 : 대저2동 관광지와 제주도 관광코스 개발



# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

- 1. 프로젝트 개괄
- 2. 현상 분석
- 3. 제공 데이터 파악
- 4. 프로젝트의 목적·목표
- 5. 프로젝트 일정

## 4. 프로젝트의 목적 · 목표 1) 연관 분석(Association Rules) – 문제 적용

### ▶ 교통의 혼잡, 안내 표지판 부족의 불편사항은 관광 코스를 찾아내면 해결할 수 있음

- 외래 관광객들이 자주 여행 다니는 코스를 발견하면 그 코스의 대중교통 외국어 안내 강화, 대중교통 도움 서비스 확대, 안내 표지판 설치 등의 방법을 적용할 수 있다.
- 아직 만들어지지 않은 외래 관광객들을 위한 관광 코스 개발이 가능하다.
- 데이터를 다음과 같은 변수 별로 분할하여 분석하였을 때, 좀 더 자세한 분석 결과를 도출해 낼 수 있다.

분할 분석할 변수	설명
DATE (날짜)	2013.12.01~2014.02.28 중 에서 크리스마스, 신정, 설 연휴에서의 분석만 따로 분석하여, 특별 마케팅을 진행 가능
COUNTRY_CODE (국가 코드)	국가별로 분석을 진행하여 국가별 주요 여행 코스에 따라, 해당 외국어 안내를 강화할 수 있음
ZIP (Call 발생지역 우편번호)	서울에 많이 집중될 가능성이 있으므로, 각 광역자치단체별로 분석하면 보다 자세한 관광 코스를 도출해 낼 수 있음



# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

- 1. 프로젝트 개괄
- 2. 현상 분석
- 3. 제공 데이터 파악
- 4. 프로젝트의 목적·목표
- 5. 프로젝트 일정

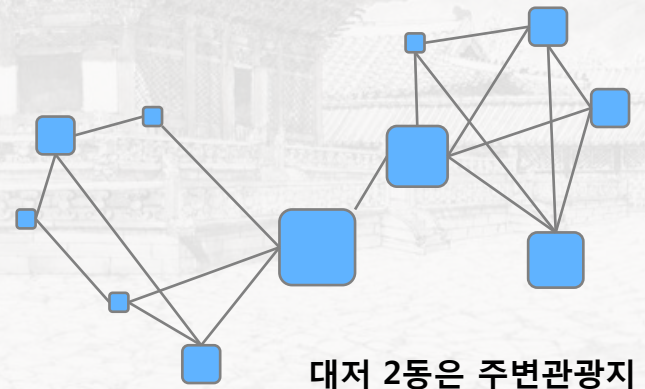
## 4. 프로젝트의 목적 · 목표 2)사회 연결망(SNA) - 설명 및 예시

▶사회 연결망 분석을 통해 대표성이 있는 장소, 관광지 그룹을 이어주는 장소 등을 파악 가능

### •사회 연결망(SNA)란?

사회학에서 개인, 집단, 사회의 관계를 네트워크로 파악하는 개념이다. 즉 개인 또는 집단이 네트워크의 하나의 노드(node)이며, 사회연결망은 이 각 노드들 간의 상호의존적인 관계(link)에 의해 만들어지는 사회적 관계 구조를 말한다.<sup>7)</sup>

- 노드(node) : ZIP (Call 발생 기지국 우편번호)
- 관계(link) : 한 관광객이 통화 발생한 장소들 집합(중복제거)
- 가상 결과 : 618807(부산광역시 강서구 대저2동)의 Betweenness Centrality가 가장 높다.
- 해석 : 부산광역시 강서구 대저2동이 다른 관광지를 이어주는 장소로 고려된다.
- 가능한 정책 : 대저2동 관광지에 숙소 수를 비교하여 부족할 경우 교통유발부담금을 줄여 부담을 줄여준다.
- 가능한 마케팅 : 대저2동에 관광 광고를 집중적으로 한다.



대저 2동은 주변관광지 그룹을 이어주는 교각역할!

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

- 1. 프로젝트 개괄
- 2. 현상 분석
- 3. 제공 데이터 파악
- 4. 프로젝트의 목적·목표
- 5. 프로젝트 일정

## 4. 프로젝트의 목적 · 목표

### 2)사회 연결망(SNA) – 문제 적용

#### ▶ 주요 중심성(Centrality)가 높은 지역 중 현저히 숙박업소가 부족한 지역 선정 가능

- 문화체육관광부에 따르면 외래 관광객들의 증가에 대비하여 숙박시설을 늘리기를 바라고 있다.
- 단순한 위치 선정이 아니라, **사회 연결망 분석을 통해 합리적인 장소**를 추천하고자 한다.
- 연관분석과 같이, 데이터를 '날짜, 국가코드, Call발생 우편번호'와 같은 변수 별로 분할하여 분석하였을 때, 좀 더 다양한 분석 결과를 도출해 낼 수 있다.
- 주요 중심성(Centrality)에 대한 해석은 다음과 같다.

Centrality	설명
In-Degree	높을 수록 관광 코스의 목적지로 되는 경우가 높은 장소
Out-Degree	높을 수록 관광 코스의 출발지로 되는 경우가 높은 장소 -관광 정보를 주는 책자 등을 배정하는 것이 좋다고 예측
Betweenness	높을 수록 관광지 그룹을 잇는 다리역할을 하는 장소 -관광 장소의 교각 지역이기에 숙박업소를 늘리는 것이 적합함
Closeness	높을 수록 대한민국 관광지를 대표하는 장소 -한국 전통/지역 문화를 알리는 행사장소로 적합함

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

- 1. 프로젝트 개괄
- 2. 현상 분석
- 3. 제공 데이터 파악
- 4. 프로젝트의 목적·목표
- 5. 프로젝트 일정

## 4. 프로젝트의 목적 · 목표 3) 기존 분석과의 비교 및 분석 적합성 평가

### ▶ 기존의 분석에서 진행하지 못한 실증적 데이터 반영

• 비교할 기본 분석 자료는 문화체육관광부가 시행하는 '2013 외래관광객 실태조사' 로 정하였다.

기존 분석	본 프로젝트 분석
<ul style="list-style-type: none"> <li>■ 매월 1000명 가량씩 오프라인으로 면접원이 직접 응답자에게 설문지를 배포하여 응답을 얻는 형식</li> <li>→ 면접원을 선발, 교육하는 비용 등이 추가로 발생</li> <li>→ 연간 1,200만 명, 향후 이보다 더 많은 수의 외래관광객을 기대하는 시점에서 매우 비효율적임</li> <li>■ 설문조사 방식은 여전히 실증적 데이터를 얻는 데에 한계가 있음</li> <li>■ 기존 분석으로는 정성적인 접근과 기초적인 정량적 접근만이 가능함</li> </ul>	<ul style="list-style-type: none"> <li>■ SKT 제공의 약 170만 Roaming Data를 기반으로 함</li> <li>■ 선정한 주제(외래관광객 만족도 증대를 위한 현재 관광 사업의 문제 해결)는 단순 설문조사 결과에서 나아가 실증적 빅 데이터를 반영하는 데 기존 분석 방법과 차별화 (예 : 단순히 전국적으로 숙박업을 늘리자는 것이 아닌 빅 데이터에 기반, 가장 적합한 위치부터 늘리자는 데에서 의의)</li> <li>■ 세부 목표(국적별 중요 관광지 코스 파악을 통한 교통 혼잡, 안내 표지판 부족 등의 문제 해결/ 대표성 있는 장소 파악을 통한 숙박업소 부족 문제 해결)의 적합성</li> <li>→ 기존 분석으로 이를 파악하려면 훨씬 더 많은 고객에게 정보를 받아야 하지만 이는 현실적으로 어렵고 빅 데이터 분석을 하면 간단한 통계적 기법으로 쉽게 패턴 파악 가능(보다 더 현실적, 경제적이고 과학적임)</li> </ul>

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

1. 프로젝트 개괄 2. 현상 분석 3. 제공 데이터 파악 4. 프로젝트의 목적·목표 5. 프로젝트 일정

## 4. 프로젝트의 목적 · 목표 4) 요약

▶ 큰 목표는 주어진 빅 데이터 이용하여, 현황 조사에서 발견한 관광 사업의 문제점을 해결하고자 함에 있음

세부 목표	방법론	분석 적합성	분석 전략	분석 활용방안 (해결 가능 문제)
관광지 이동 경로 연관 분석	연관 분석 (AR)	기존 단순 통계 분석이 못하는 관광지 여행(관광 코스)의 흐름을 발견해 낼 수 있음	(국가별로 데이터 분할 →) 전처리를 통해 ID(고객)와 ITEM(장소) 변수를 정제 → 동일데이터 삭제 → SAS 9.4로 AR진행 → 결과 해석	-주요 관광 코스에 대해 대중 교통 지원, 표지판 설치 (국가별로 진행 시, 대중 교통에 해당 외국어 방송 지원)
관광지 연결망 파악 및 주요장소 파악	사회 연결망 (SNA)	기존 단순 통계 분석이 못하는 관광지 그룹을 잇는 다리 역할을 하는 장소 (Betweenness Centrality)와 지역을 대표하는 장소 (Closeness centrality)를 발견해 낼 수 있음	전처리를 통해 고객과 장소 변수를 정제 → 동일데이터 삭제 → Node(장소) 와 Link(한 고객이 전화한 지역;순서 반영)로 정제 → UCINET 6로 SNA 진행 → 다양한 중심성(Centrality) 위주로 해석	-중저가 숙박 시설 현황과 비교 분석하여 중저가 숙박 시설 부족 지역 발견 -Degree Centrality가 높은 지역의 경우 관광 정보 책자 등을 제공

\*추가 필요 데이터 : 숙박 시설 현황(SNA 결과와 비교 분석), 우편 번호 정보(우편 번호-장소명 변환)

# 설문 조사로 관광 정책을 제시하는 시대는 떠났다!

1. 프로젝트 개괄 2. 현상 분석 3. 제공 데이터 파악 4. 프로젝트의 목적·목표 5. 프로젝트 일정

## 5. 프로젝트 일정

프로젝트 수행 과제		프로젝트 일정							
		~4/20	~4/24	~4/27	~4/30	~5/11	~6/18	~6/1	~6/10
기획 부문	분석 과제 파악 및 선행연구 탐색	■	■						
	데이터 수령 신청 및 수령		■						
	분석 가을 목적·목표 설정		■	■					
	기획 부문 보고서 작성			■	■				
분석 부문	데이터 전처리					■			
	연관 분석 진행					■	■		
	사회 연결망 진행					■	■		
	정책/마케팅 전략 수립						■	■	
	추가 분석 및 정책/마케팅 전략 수립							■	■
	분석 부문 보고서 작성						■	■	■

## 6. 참고문헌

---

- 1) 문화체육관광부. (2014.2). *2013 외래관광객 실태조사*.
- 2) 송주연. (2014). "제주도는 중국인, 울산엔 노르웨이인 관광상품 뜬다?". The bell. Retrieved from [http://thebell.co.kr/front/free/contents/news/article\\_view.asp?key=201404090100015370000942&svccode=03&page=1](http://thebell.co.kr/front/free/contents/news/article_view.asp?key=201404090100015370000942&svccode=03&page=1). (2014.04.24. 열람)
- 3) 한국데이터베이스진흥원, 『데이터 분석 전문가 가이드』(한국데이터베이스진흥원, 2014).
- 4) 변진경. (2013). "호텔의 호시절, 얼마나 갈까?". 시사iN. Retrieved from <http://www.sisainlive.com/news/quickViewArticleView.html?idxno=17474>. (2014.04.24. 열람)
- 5) 박태정. (2014). "중저가 관광호텔 필요한데 특급만 '줄줄'. 파이낸셜뉴스. Retrieved from [http://www.fnnews.com/view?ra=Sent1301m\\_View&corp=fnnews&arcid=14040306000867&cDateYear=2014&cDateMonth=04&cDateDay=03](http://www.fnnews.com/view?ra=Sent1301m_View&corp=fnnews&arcid=14040306000867&cDateYear=2014&cDateMonth=04&cDateDay=03). (2014.04.24 열람)
- 6) Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Mining association rules between sets of items in large databases". Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93. p. 207. doi:10.1145/170035.170072. ISBN 0897915925.
- 7) D'Andrea, Alessia et al. (2009). "An Overview of Methods for Virtual Social Network Analysis". In Abraham, Ajith et al. Computational Social Network Analysis: Trends, Tools and Research Advances. Springer. p. 8. ISBN 978-1-84882-228-3.