

유튜브 추천 데이터로 만들어본, 유튜브 채널간 관계지도

샌드박스네트워크 | Data Lab | 노성산

2019. 10. 16



샌드박스에서 CX & Data Lab 을 맡고 있습니다.

인문학과 공학 사이에서 고생하는 중

- 유튜브 크리에이터 출신
- 컴퓨터 공학 + MBA
- 3번째 스타트업 (엑싯은 언제 해보나?)

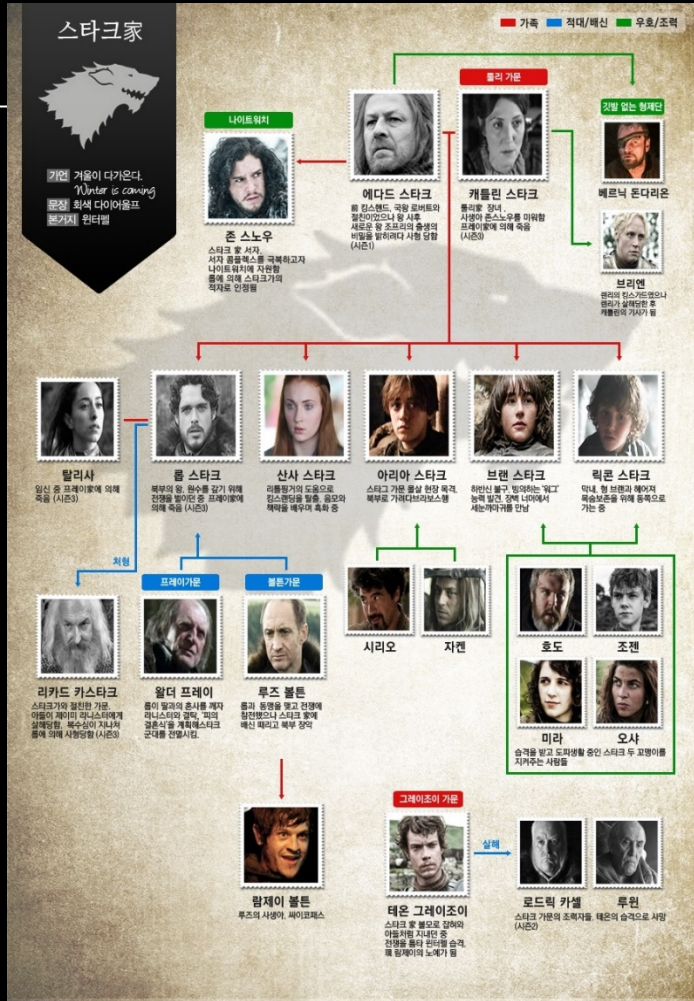


엇? 그럼 사이드 프로젝트가 아니라
본업 아닌가요?



유튜브 채널 관계도에 사용된 도구들

- Python
- YouTube Data API
- MySQL
- Excel
- NodeXL Basic (무료)



재료 파악이 제일 중요하다

SOSCON2019

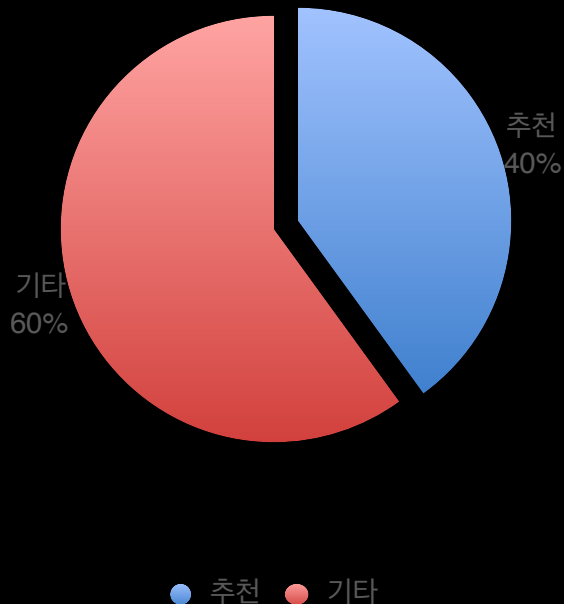
데이터를 통해 무언가 하고 싶다면
내가 (가진 or 가질) Raw Data 가 무엇인지 명확하게 알아야 한다.



우리는 이미 유튜브의 노예

1. 관심사 기반의 추천

2. 키워드 기반의 추천



추천 데이터를 가져오는 방법

1. Web Crawling

검증용

2. YouTube Data API

수집용

<https://developers.google.com/youtube/v3/docs/search>

가장 큰 장점은 옵션이 많은 것이 아니라

“검증이 가능하다는 것”



API 를 사용하여 특정 영상의 추천 데이터 가져오기

SOSCON2019

```
youtube = build(name, version, key)
output = youtube.search().list(
    part="snippet",
    relatedToVideoId=video_id,
    maxResults=num_of_results,
    type="video",
    regionCode="kr"
).execute()
```



난이도 하, 누워서 핸드폰 보기 수준

문제는?

보검 TV

3.54M subscribers • 8,462 videos



영상 별로 추천 데이터를 뽑아보았다

SOSCON2019



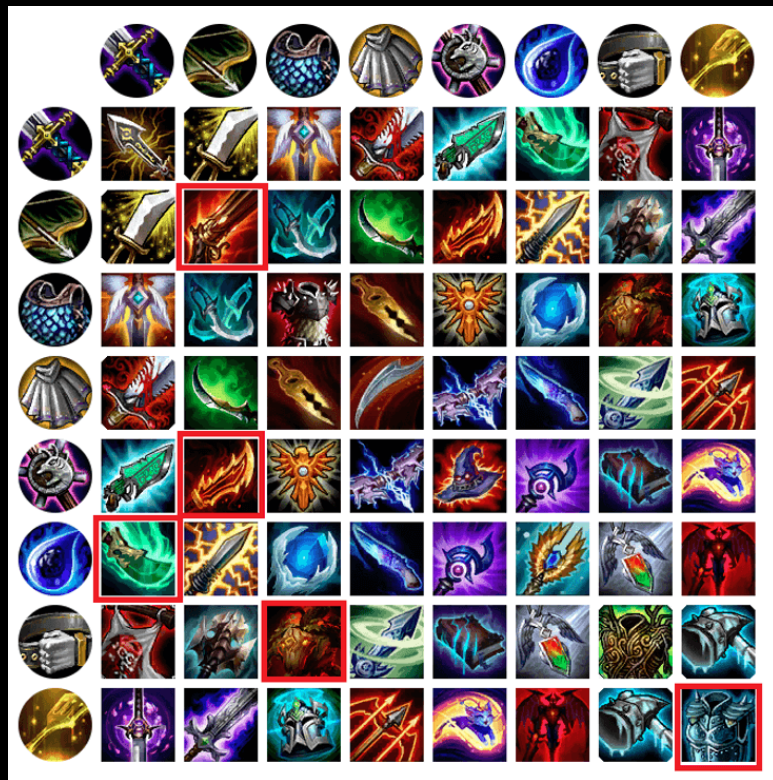
결국 이 프로젝트의 핵심은?

채널을 대표하는 영상을 어떤 기준으로 선별할 것인가?

몇 개의 추천 데이터가 있어야 실제 데이터와 95% 이상 유사할까?

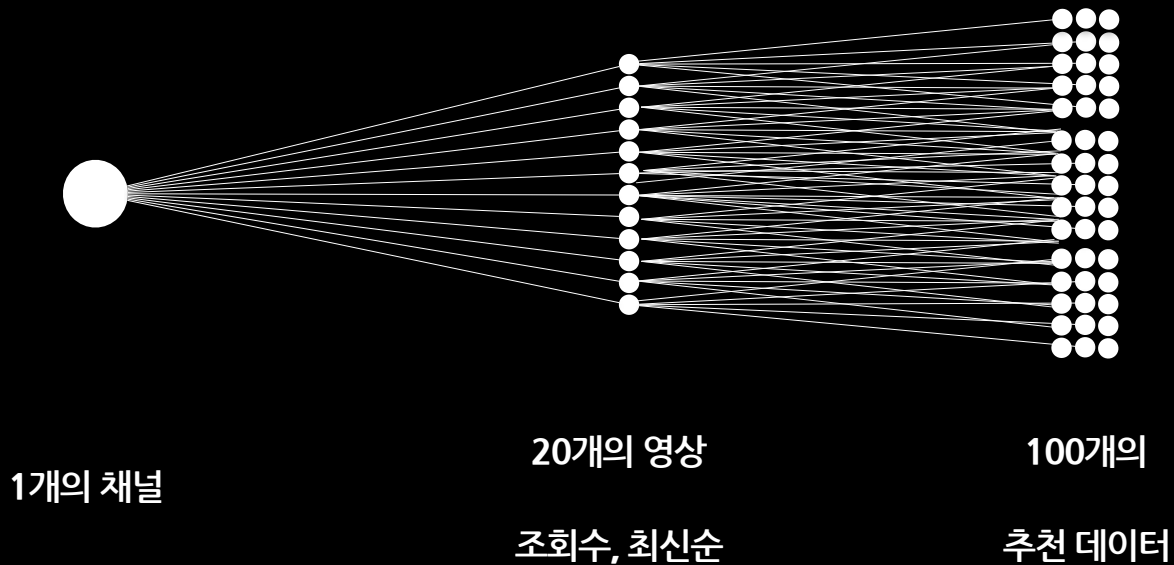
원하는 결과가 나올 때까지

다양한 조합을 시도



그나마 가장 정확한 결과를 보여준...

SOSCON2019



100개의 추천 데이터를 압축해보자

100개에서 등장빈도가 높은 5개의 데이터만 추출

Collections 모듈의 Counter 함수를 쓰면 편리

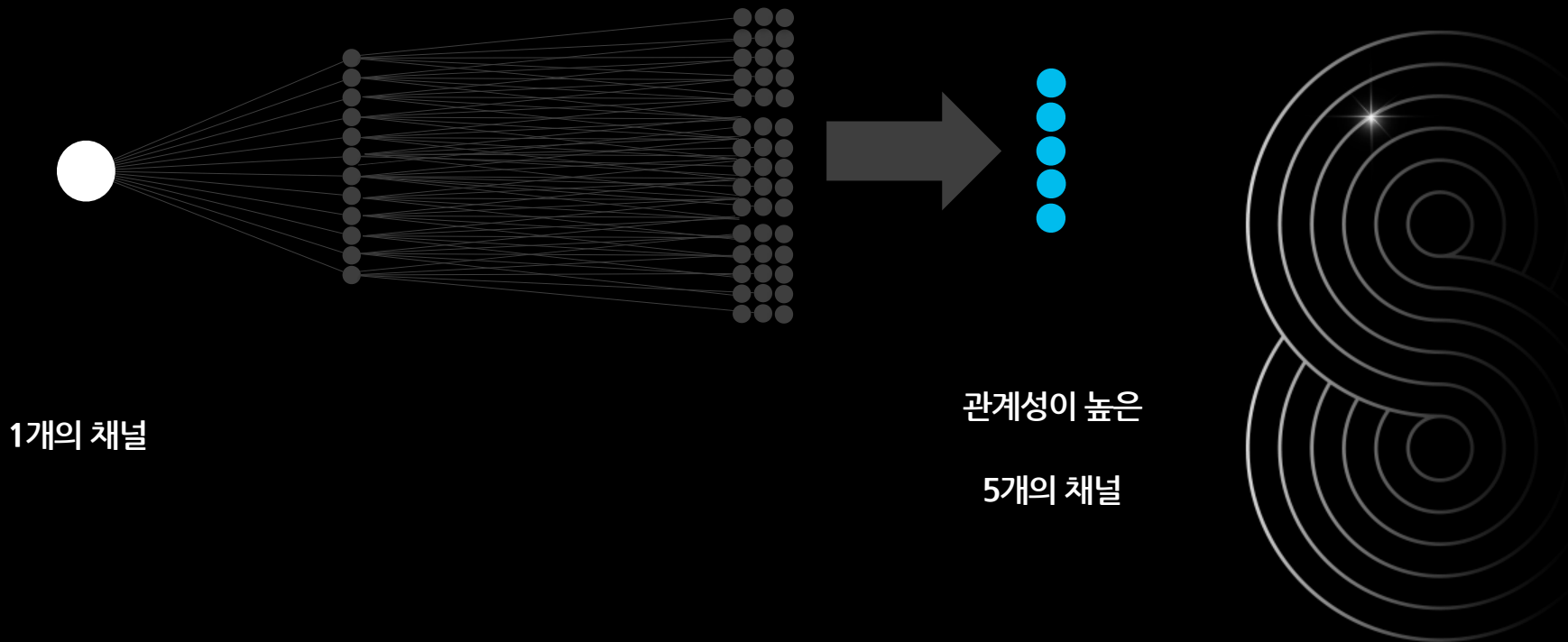
```
import collections
```

```
lst = [ 'a', 'b', 'b', 'c', 'c', 'c', 'd' ]
```

```
c = collections.Counter ( lst )
```

```
print ( c.most_common ( 2 ) )
```





실제 결과는?

SOSCON2019

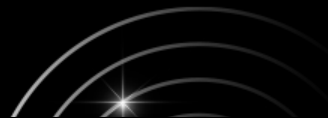


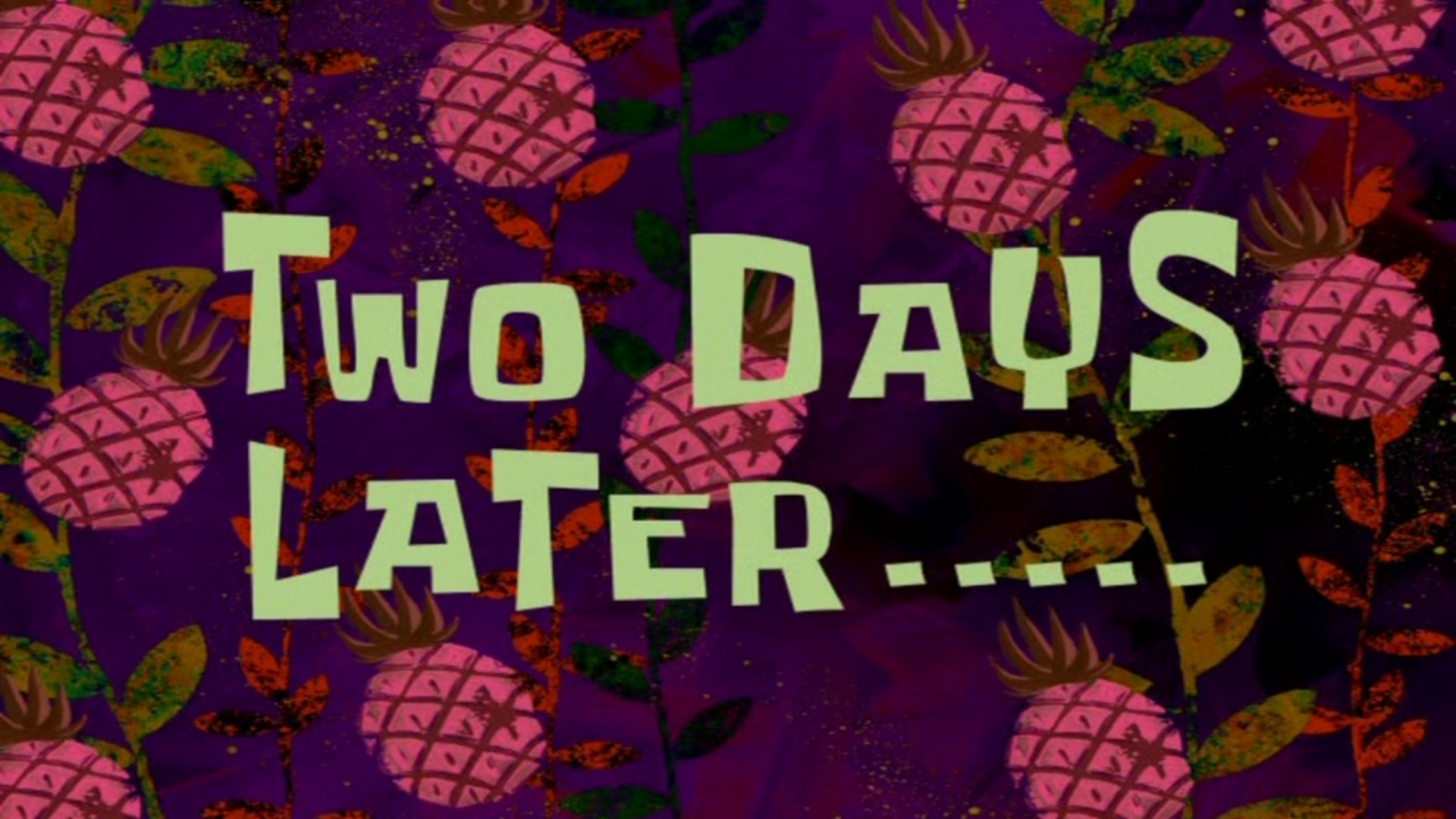
for i in range(2500)

SOSCON2019

국내를 타겟으로 하는 구독자 10만명 이상 채널

2500개이 초킹 데이터르 가가 2개씩 쓰지

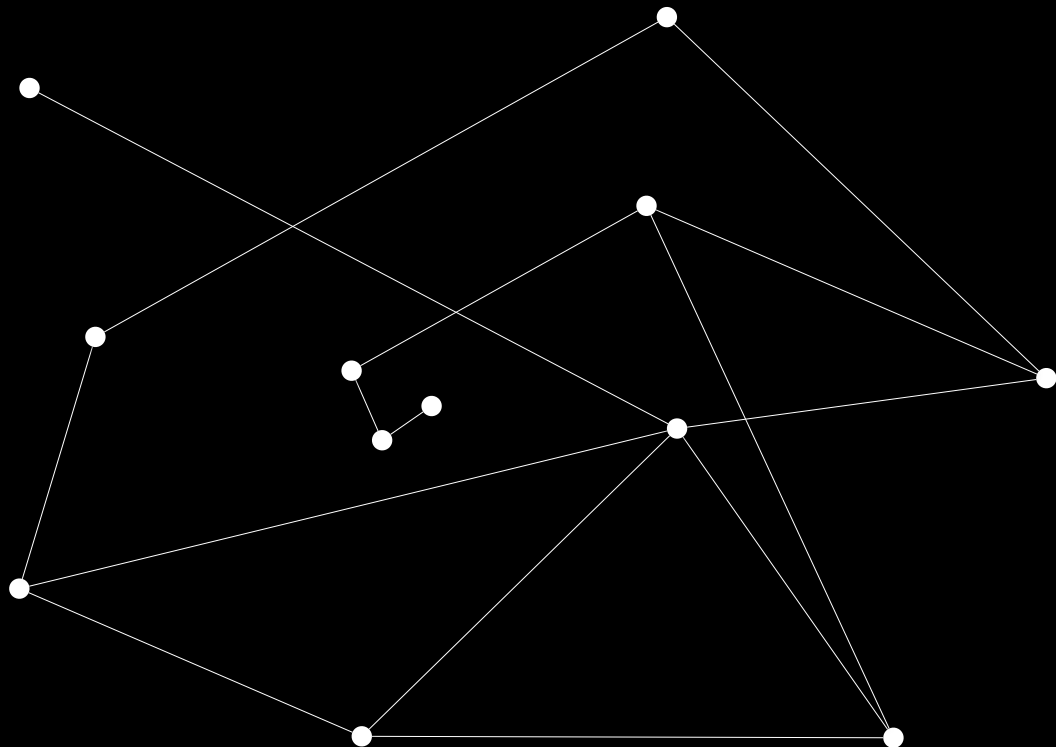




**TWO DAYS
LATER - - - -**

수집된 데이터를 NodeXL 에 그대로 부어보았다

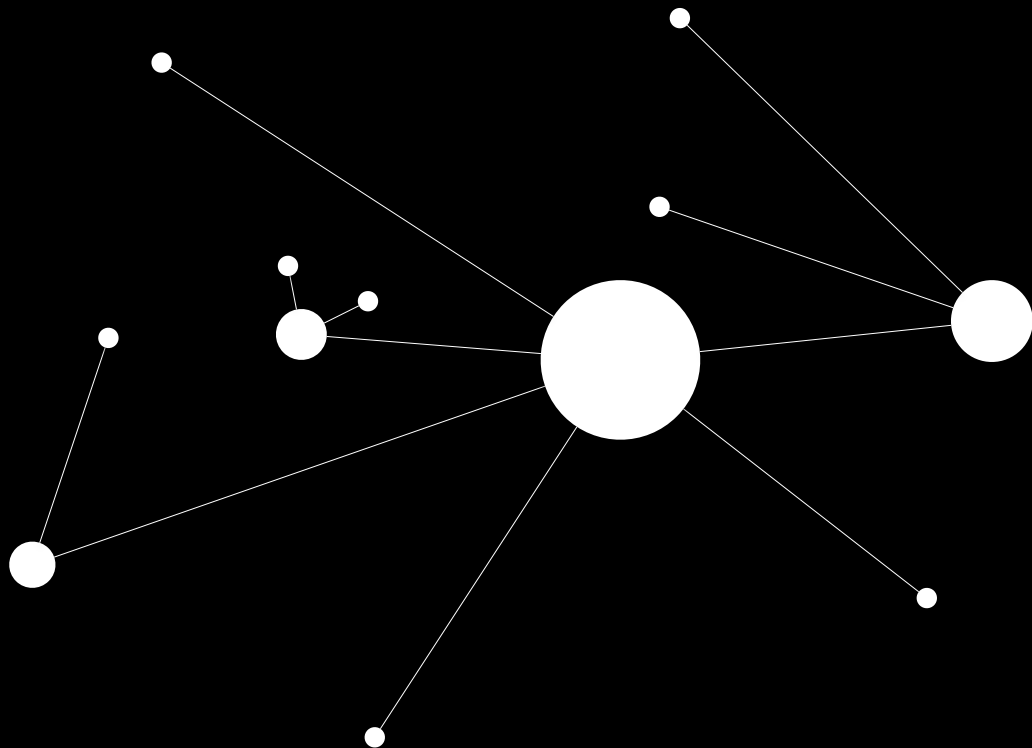
SOSCON2019



아무것도 하지 않고
데이터를 넣었을 때 결과

NodeXL 에 맞게 바꾸어 부어보았다

SOSCON2019



내가 고생하는 만큼
보는 사람이 편해진다

1. 자주 등장하는 채널들에 대한 가중치 표현

대중적이거나 장르 대표 채널들이

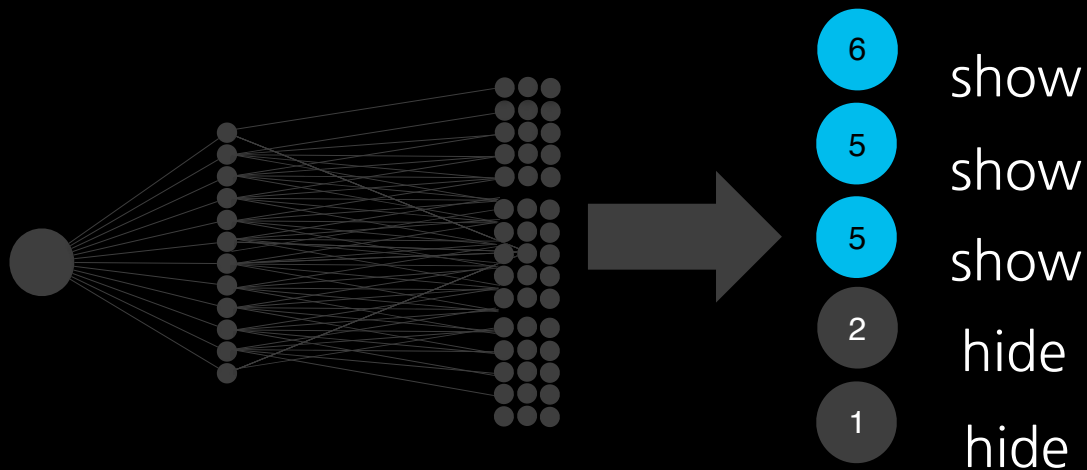
추천 채널로 자주 등장

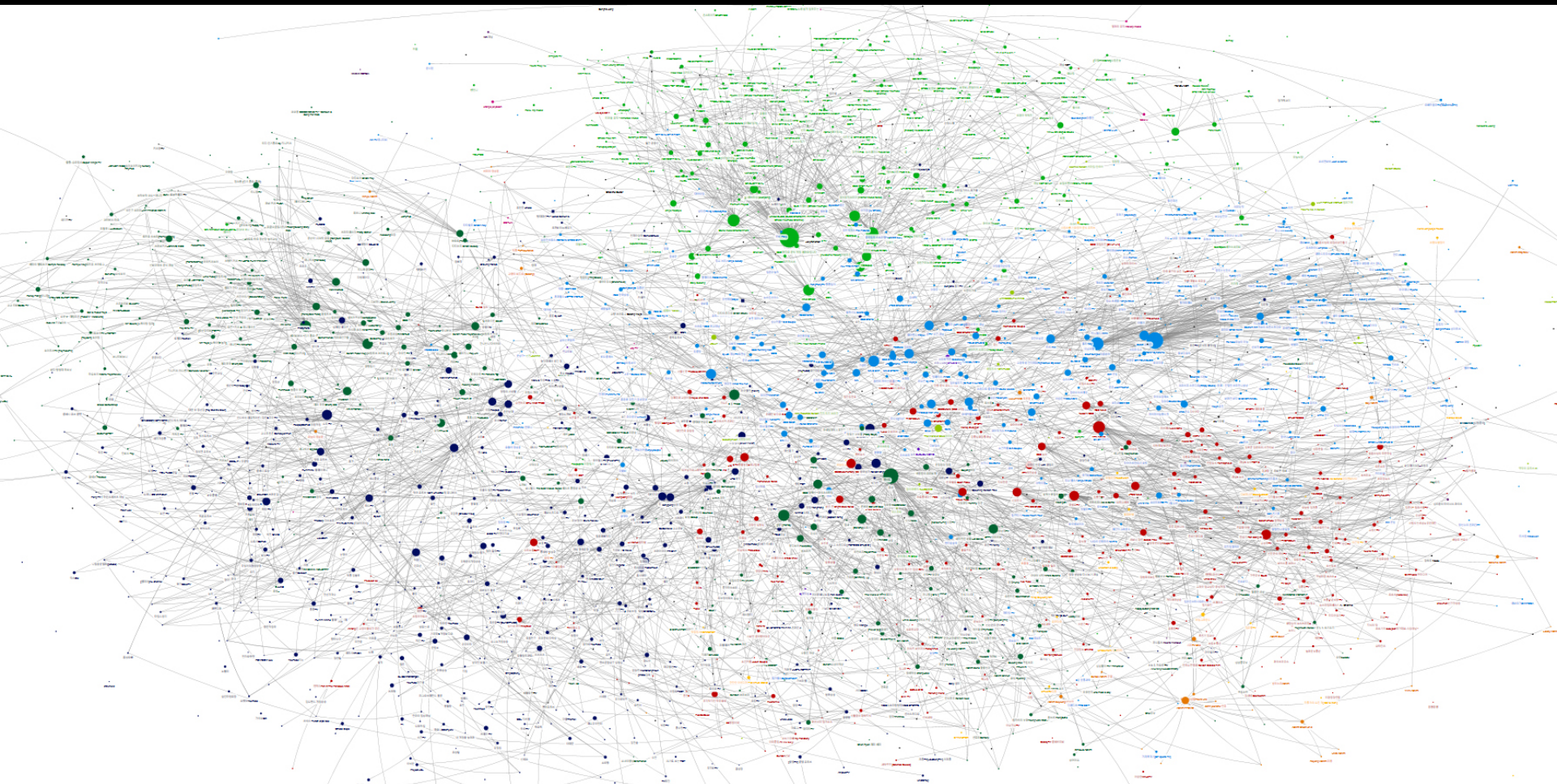


cid	network	weight
UCPZ2XXfOWDXi7dGlQsOt5IQ	UCweOkPb1wVVH0Q0Tlj4a5Pw	7
UCB4HIC-s6wpVvT8o1D1n7aQ	UCweOkPb1wVVH0Q0Tlj4a5Pw	6
UCJ_sOIMLKpEH39gESJCPmlw	UCweOkPb1wVVH0Q0Tlj4a5Pw	6
UCgxJhO66X-95iZGe4a_ku3Q	UCweOkPb1wVVH0Q0Tlj4a5Pw	6
UC8HNshpReWjQv1WpwzhPHjA	UCweOkPb1wVVH0Q0Tlj4a5Pw	5
UCguxVQUtEIa6hcgZdhj_yA	UCweOkPb1wVVH0Q0Tlj4a5Pw	5
UC0qlzFa7YUtWW07Dqr-9XA	UCweOkPb1wVVH0Q0Tlj4a5Pw	5
UCN7DYShSVVL3Uvy3LQszhUg	UCweOkPb1wVVH0Q0Tlj4a5Pw	5
UCrhgVHA_uAT9BoGV9qRKvg	UCweOkPb1wVVH0Q0Tlj4a5Pw	5
UC8QN00_H7Vg3NIIdO-6_E8aA	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UCzd-QY2q6hFyNb_bn9vtWQw	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UCk11rbOrsyPfBuHNfnLPm-Q	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UCtwDdyMko-1-rJ8yPyrOTThDA	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UC4UqChiuMVMtmerrSayWIQA	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UCji4MHZnUaijm2s3tpHOLJg	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UCZECMCJJJE47ei170XCXgdA	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UCnOTemIK0FqCe2FHbr9mWFW	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UCRDd3x33kf0IW6g2MRUkRw	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UCS7W22-coWXOnlyN97S2pvg	UCweOkPb1wVVH0Q0Tlj4a5Pw	4
UCTE8ueKNA54RcO6MgXif2Tg	UCweOkPb1wVVH0Q0Tlj4a5Pw	4

2. 관계성 지수가 높은 채널의 edge 만 표현

→ Edge 의 visibility 에 반영





The graph illustrates a dense network of relationships within the Korean media landscape. Key clusters include:

- SBS Network:** A large cluster centered around SBS Catch, SBS ENTER PLAY, and SBS NOW, with numerous smaller nodes representing specific programs or affiliates.
- EBS Network:** A significant hub around EBS Culture (EBS 교양) and EBS Documentary (EBS 다큐), connected to various educational and documentary-related entities.
- tvN Network:** A group of nodes including tvN D ENT, Travel Tube, and Maangchi, indicating a strong interconnectedness within this channel's ecosystem.
- Independent Producers/Creators:** Numerous smaller nodes represent individual content creators and independent production companies, often linked to larger networks.

The overall structure suggests a highly competitive and collaborative environment where media entities frequently cross-promote or share resources across different platforms and genres.

Clustering 결과만 요약하면

- 01 대중 : 방송미디어, 탑크리에이터, 먹방
- 02 학생 : 게임, 예능형 크리에이터, 영화 리뷰
- 03 뮤직 : 아이돌, 음악방송국
- 04 코미디 : 코미디, 축구, 정치
- 05 키즈 : 2세~8세
- 06 뷰티 : 메이크업
- 07 펫 : 강아지, 고양이, 파충류
- 08 영어공부
- 09 인디뮤직
- 10 공포



더 자세하게 보고 싶으신 분들은

bit.ly/youtube_map

무언가 함께 해보고 싶으시다면?

sungsan@sandbox.co.kr

SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019



THANK YOU

SOSCON 2019

SAMSUNG OPEN SOURCE CONFERENCE 2019

