

# 대본공작소: SNA

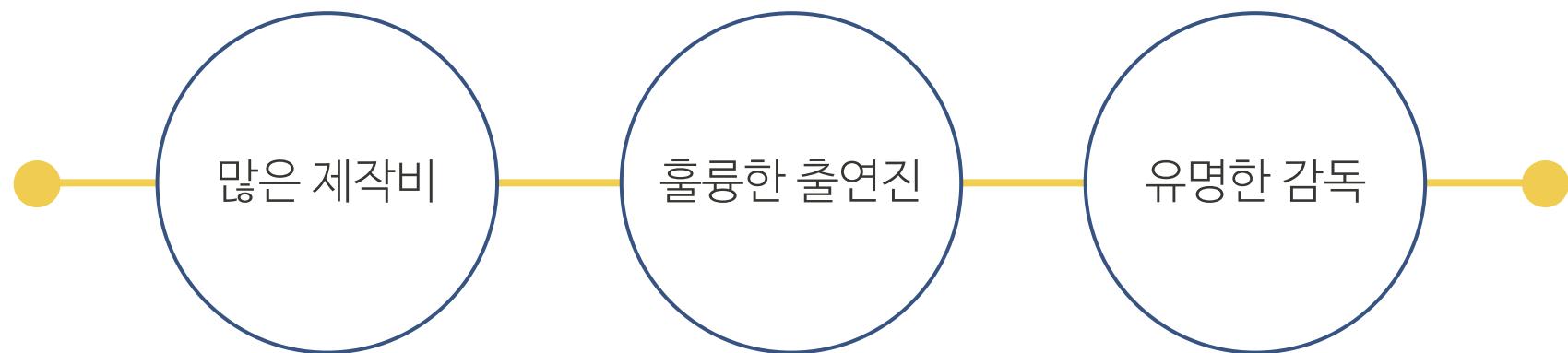
이야기의 구조가 실제로 영화의 흥행 여부에 영향을 미칠까?

김도영 | 김연진 | 박다정 | 이미소 | 황소희

## INTRODUCTION

어떤 영화가 흥할까?

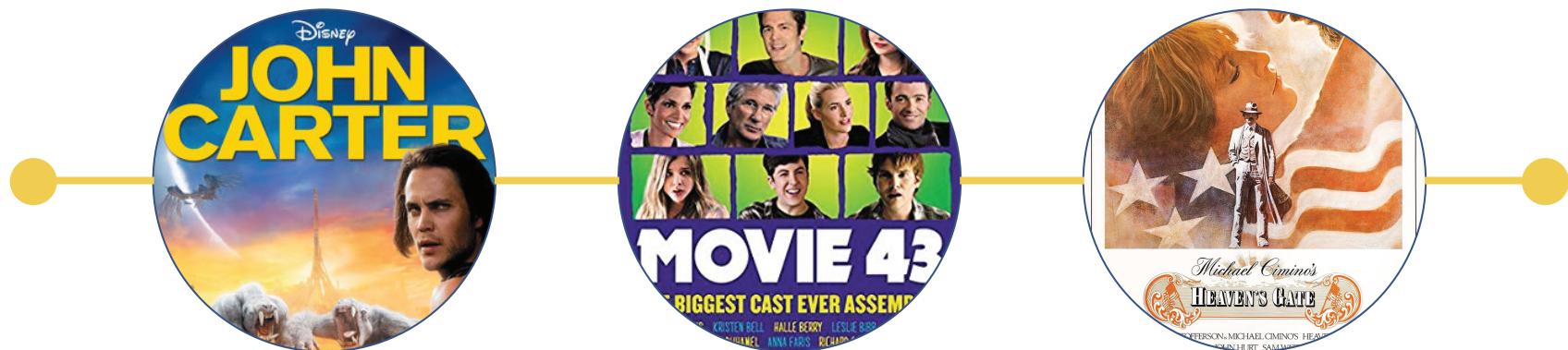
개봉전 흥행이 기대되는 영화?



## INTRODUCTION

어떤 영화가 흥할까?

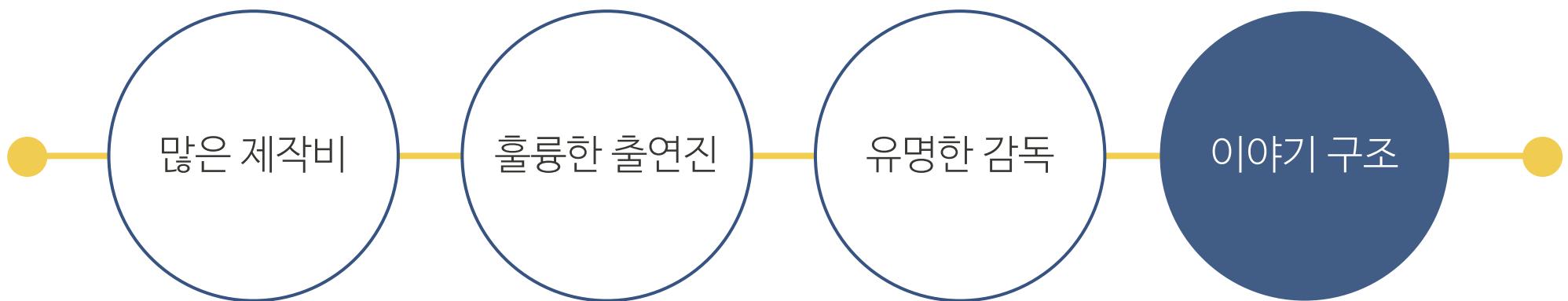
### 개봉전 흥행이 기대되는 영화?



## INTRODUCTION

어떤 영화가 흥할까?

혹시 이야기 구조에 차이가 있는 것은 아닐까?



# INTRODUCTION

어떤 영화가 흥할까?

장르별 흥행 성공 영화와 흥행 실패 영화 12작 선정

| 흥행 성공 영화            | 장르       | 흥행 실패 영화              | 장르            |
|---------------------|----------|-----------------------|---------------|
| Get Out             | 공포, 스릴러  | Annihilation          | 공포, 액션        |
| Moonlight           | 드라마      | Kill Your Darlings    | 드라마, 로맨스      |
| King's Speech       | 드라마      | The Hollars           | 드라마, 로맨스, 코미디 |
| 12 Years a Slave    | 드라마      | Loving                | 드라마, 스릴러      |
| Midnight in Paris   | 코미디, 멜로  | The Kings of Summer   | 드라마, 코미디      |
| Birdman             | 코미디, 드라마 | Infinitely Polar Bear | 드라마, 코미디      |
| Starwars            | 액션, SF   | Almost Famous         | 드라마, 코미디      |
| Avatar              | 액션, SF   | Speed Racer           | 액션, 가족        |
| Avengers            | 액션, 판타지  | Serenity              | 액션, SF        |
| Argo                | 드라마, 스릴러 | 47 Ronin              | 액션, 판타지       |
| Slumdog Millionaire | 범죄, 드라마  | Neon Demon            | 드라마, 스릴러      |
| Spotlight           | 드라마, 스릴러 | A Most Violent Year   | 드라마, 스릴러      |

## INTRODUCTION

어떤 영화가 흥할까?

장르별 흥행 성공 영화와 흥행 실패 영화 12작 선정

| 흥행 성공 영화            | 제작비     | 매출         | 흥행 실패 영화              | 제작비     | 매출      |
|---------------------|---------|------------|-----------------------|---------|---------|
| Get Out             | 45억 원   | 2550억 원    | Annihilation          | 470억 원  | 429억 원  |
| Moonlight           | 40억 원   | 650억 원     | Kill Your Darlings    | 57억 원   | 17억 원   |
| King's Speech       | 150억 원  | 4140억 원    | The Hollars           | 38억 원   | 11억 원   |
| 12 Years a Slave    | 200억 원  | 1877억 원    | Loving                | 90억 원   | 92억 원   |
| Midnight in Paris   | 170억 원  | 1510억 원    | The Kings of Summer   | 15억 원   | 14억 원   |
| Birdman             | 180억 원  | 1030억 원    | Infinitely Polar Bear | 67억 원   | 18억 원   |
| Starwars            | 3060억 원 | 2조 680억 원  | Almost Famous         | 600억 원  | 474억 원  |
| Avatar              | 2900억 원 | 2.8조 원     | Speed Racer           | 1262억 원 | 939억 원  |
| Avengers            | 2200억 원 | 1조 5119억 원 | Serenity              | 390억 원  | 388억 원  |
| Argo                | 445억 원  | 2323억 원    | 47 Ronin              | 1750억 원 | 1517억 원 |
| Slumdog Millionaire | 150억 원  | 3779억 원    | Neon Demon            | 70억 원   | 34억 원   |
| Spotlight           | 200억 원  | 880억 원     | A Most Violent Year   | 200억 원  | 120억 원  |

## INTRODUCTION

어떤 영화가 흥할까?

장르별 흥행 성공 영화와 흥행 실패 영화 12작 선정

| 흥행 성공 영화            | 제작비     | 매출         | 흥행 실패 영화              | 제작비     | 매출      |
|---------------------|---------|------------|-----------------------|---------|---------|
| Get Out             | 45억 원   | 2550억 원    | Annihilation          | 470억 원  | 429억 원  |
| Moonlight           | 40억 원   | 650억 원     | Kill Your Darlings    | 57억 원   | 17억 원   |
| King's Speech       | 150억 원  | 4140억 원    | The Hollars           | 38억 원   | 11억 원   |
| 12 Years a Slave    | 200억 원  | 1877억 원    | Loving                | 90억 원   | 92억 원   |
| Midnight in Paris   | 170억 원  | 1510억 원    | The Kings of Summer   | 15억 원   | 14억 원   |
| Birdman             | 180억 원  | 1030억 원    | Infinitely Polar Bear | 67억 원   | 18억 원   |
| Starwars            | 3060억 원 | 2조 680억 원  | Almost Famous         | 600억 원  | 474억 원  |
| Avatar              | 2900억 원 | 2.8조 원     | Speed Racer           | 1262억 원 | 939억 원  |
| Avengers            | 2200억 원 | 1조 5119억 원 | Serenity              | 390억 원  | 388억 원  |
| Argo                | 445억 원  | 2323억 원    | 47 Ronin              | 1750억 원 | 1517억 원 |
| Slumdog Millionaire | 150억 원  | 3779억 원    | Neon Demon            | 70억 원   | 34억 원   |
| Spotlight           | 200억 원  | 880억 원     | A Most Violent Year   | 200억 원  | 120억 원  |

## INTRODUCTION

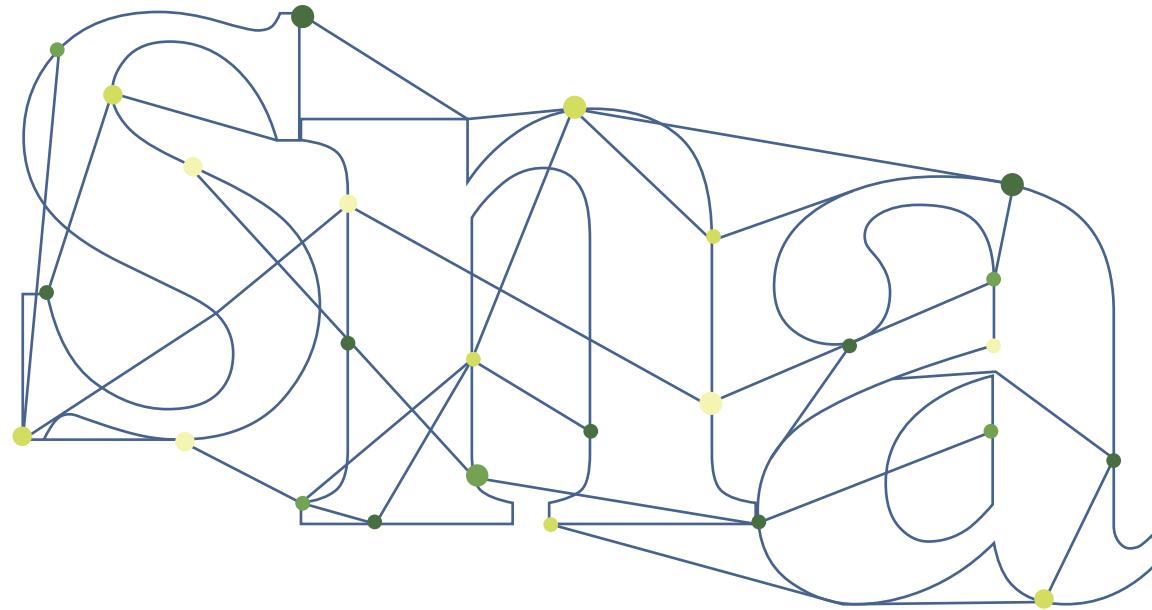
어떤 영화가 흥할까?

장르별 흥행 성공 영화와 흥행 실패 영화 12작 선정

| 흥행 성공 영화            | 제작비     | 매출         | 흥행 실패 영화              | 제작비                  | 매출      |
|---------------------|---------|------------|-----------------------|----------------------|---------|
| Get Out             | 45억 원   | 2550억 원    | Annihilation          | 470억 원               | 429억 원  |
| Moonlight           | 40억 원   | 650억 원     | Kill Your Darlings    | 57억 원                | 17억 원   |
| King's Speech       | 150억 원  | 4140억 원    | The Hollars           | 38억 원                | 11억 원   |
| 12 Years a Slave    | 200억 원  | 1877억 원    | Loving                | 90억 원                | 92억 원   |
| Midnight in Paris   | 170억 원  | 1510억 원    | The Kings of Summer   | 15억 원                | 14억 원   |
| Birdman             | 180억 원  | 1030억 원    | Infinitely Polar Bear | 매출이 제작비보다 적거나 유사한 영화 | 18억 원   |
| Starwars            | 3060억 원 | 2조 680억 원  | Almost Famous         | 474억 원               |         |
| Avatar              | 2900억 원 | 2.8조 원     | Speed Racer           | 1262억 원              | 939억 원  |
| Avengers            | 2200억 원 | 1조 5119억 원 | Serenity              | 390억 원               | 388억 원  |
| Argo                | 445억 원  | 2323억 원    | 47 Ronin              | 1750억 원              | 1517억 원 |
| Slumdog Millionaire | 150억 원  | 3779억 원    | Neon Demon            | 70억 원                | 34억 원   |
| Spotlight           | 200억 원  | 880억 원     | A Most Violent Year   | 200억 원               | 120억 원  |

## METHOD

### Social Network Analysis



사람, 그룹, 조직, 컴퓨터 및 데이터 등 객체 간의 관계 및 네트워크의 특성과 구조를 분석하고 시각화하는 첨단 분석 방법론. 1980년대 이후 수리사회학과 수학의 그래프 이론의 결합을 통해 급속히 발전되어 왔다. 최근 범죄 수사, 첨보, 조직 분석, 커뮤니케이션망 분석, 에이즈(AIDS) 확산 연구, 제약 연구 등의 분야에 활발하게 응용되고 있다.

출처: 네이버 지식백과

## METHOD

Gephi



구글, 오라클, 미 국립보건연구원 등에서 사용되는 프로그램으로 경로, 네트워크, 사회관계 등을  
분석해 그래프로 나타내는 오픈소스

출처: [세계일보] 사이버사 댓글 수사에 쓰인 '사회관계망분석' '빅데이터'는 무엇

# PREPROCESSING

## 전처리 과정

파일변환  
&인코딩

장면 구분에서 제외해야  
되는 줄 제거

scene 구분자  
모두 찾기

scene 별로  
리스트에 담기

인물 구분자  
찾기

인물들을  
리스트로 담기

ANDRE (CONT'D)  
You know this could be considered  
stalking and harassment. Plus,  
you're sitting in a deadly weapon  
so that's some felony-type shit  
right there.

No response.

ANDRE (CONT'D)  
What?!? Say something then.

ANDRE  
(weakly)  
Help

Andre passes out as the security light floods the la  
5.

The Shaw family eats while Andre lays on their front  
This time Richard, buried in his tablet, doesn't not:  
driver lift Andre's limp body and carries him to his

JOSHUA  
Tony said Mickey's face doesn't

csv 파일로 저장

각 scene에서 한 줄씩  
확인하면서 대사하는  
인물 나오면 체크  
전체 메트릭스에 인물별  
가중치 계산해서 더하기

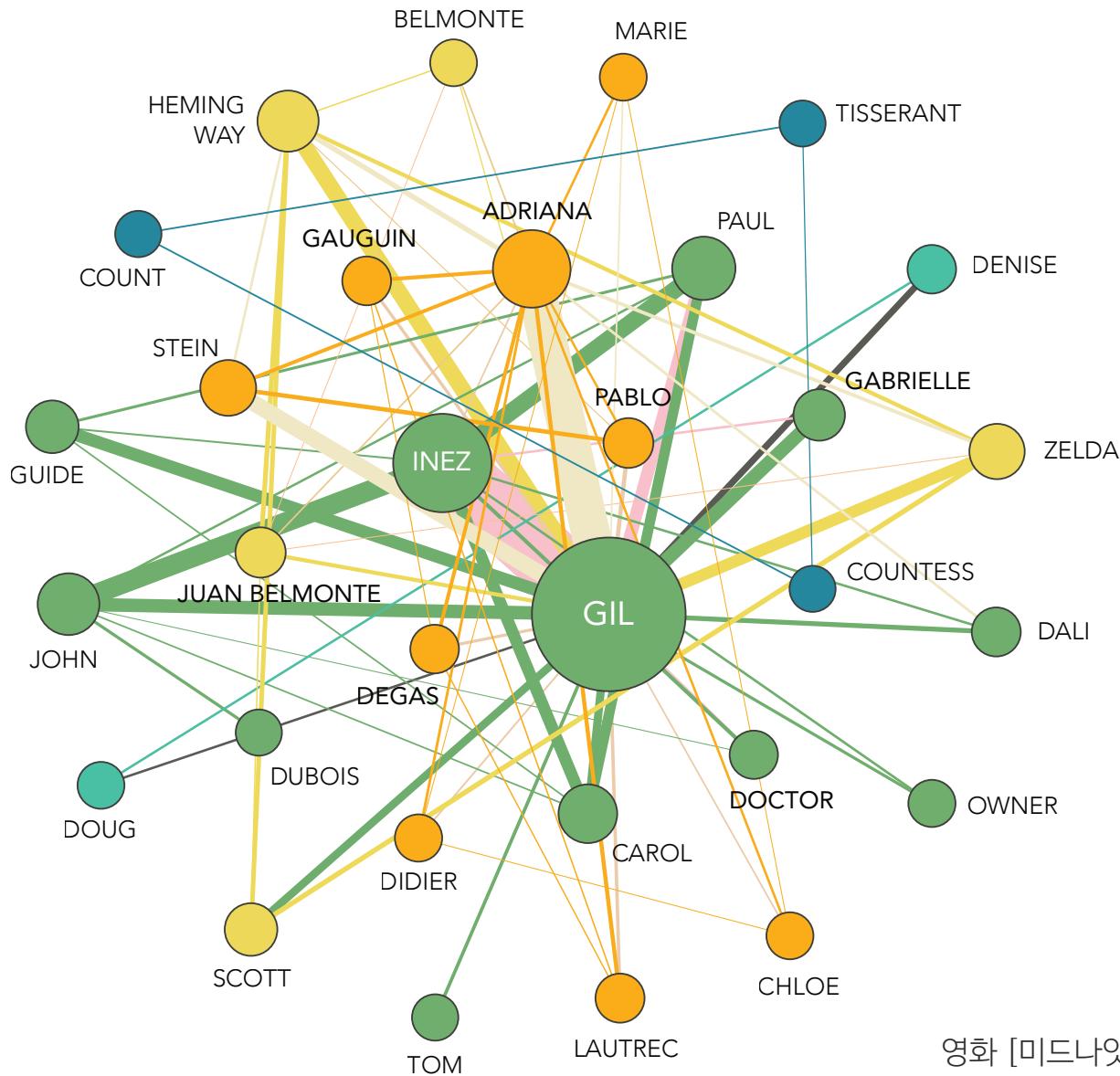
인물수 매트릭스  
만들기

| A                | B           | C           | D           | E           | F           | G           |
|------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 1 JAVED          | 0           | 0           | 6           | 0           | 0           | 14.92491617 |
| 2 PREM           | 0           | 0           | 147.3053185 | 1.837117307 | 0           | 0           |
| 4 JAMAL          | 6           | 147.3053185 | 0           | 24.27506261 | 3.432890936 | 116.0062209 |
| 5 CONSTABLE SR   | 0           | 1.837117307 | 24.27506261 | 0           | 4.566784355 | 0           |
| 6 YOUNG CONST.   | 0           | 0           | 3.432890936 | 4.566784355 | 0           | 0           |
| 7 SALIM          | 14.92491617 | 0           | 116.0062209 | 0           | 0           | 0           |
| 8 SECURITY GUARD | 0           | 1.224744871 | 5.311862178 | 0           | 0           | 0           |
| 9 PRAKASH        | 0           | 0           | 0           | 0           | 0           | 5.291502622 |
| 10 MOTHER        | 0           | 0           | 6.415560765 | 0           | 0           | 2.108185107 |
| 11 DIRECTOR      | 0           | 9.553968567 | 2.165063509 | 0           | 0           | 0           |
| 12 KRISHNA       | 0           | 0           | 2.108185107 | 0           | 0           | 1.264911064 |
| 13 LATIKA        | 8.521243446 | 8.797016118 | 65.53430032 | 0           | 0           | 31.4375044  |
| 14 NITA          | 0           | 12.89897949 | 0           | 0           | 0           | 0           |
| 15 ARVIND        | 0           | 0           | 11.85899172 | 0           | 0           | 3.338662108 |
| 16 PUNNOOSE      | 0           | 0           | 2.772113675 | 0           | 0           | 6.755551789 |
| 17 GUIDE         | 0           | 0           | 0           | 0           | 0           | 0           |
| 18 ADA           | 0           | 0           | 5.850101214 | 0           | 0           | 0           |

인물 리스트에서  
제외해야 할 단어 제거  
(인물 이름 오타, 인물 별명 등)

# BASICS

통계량 이해하기

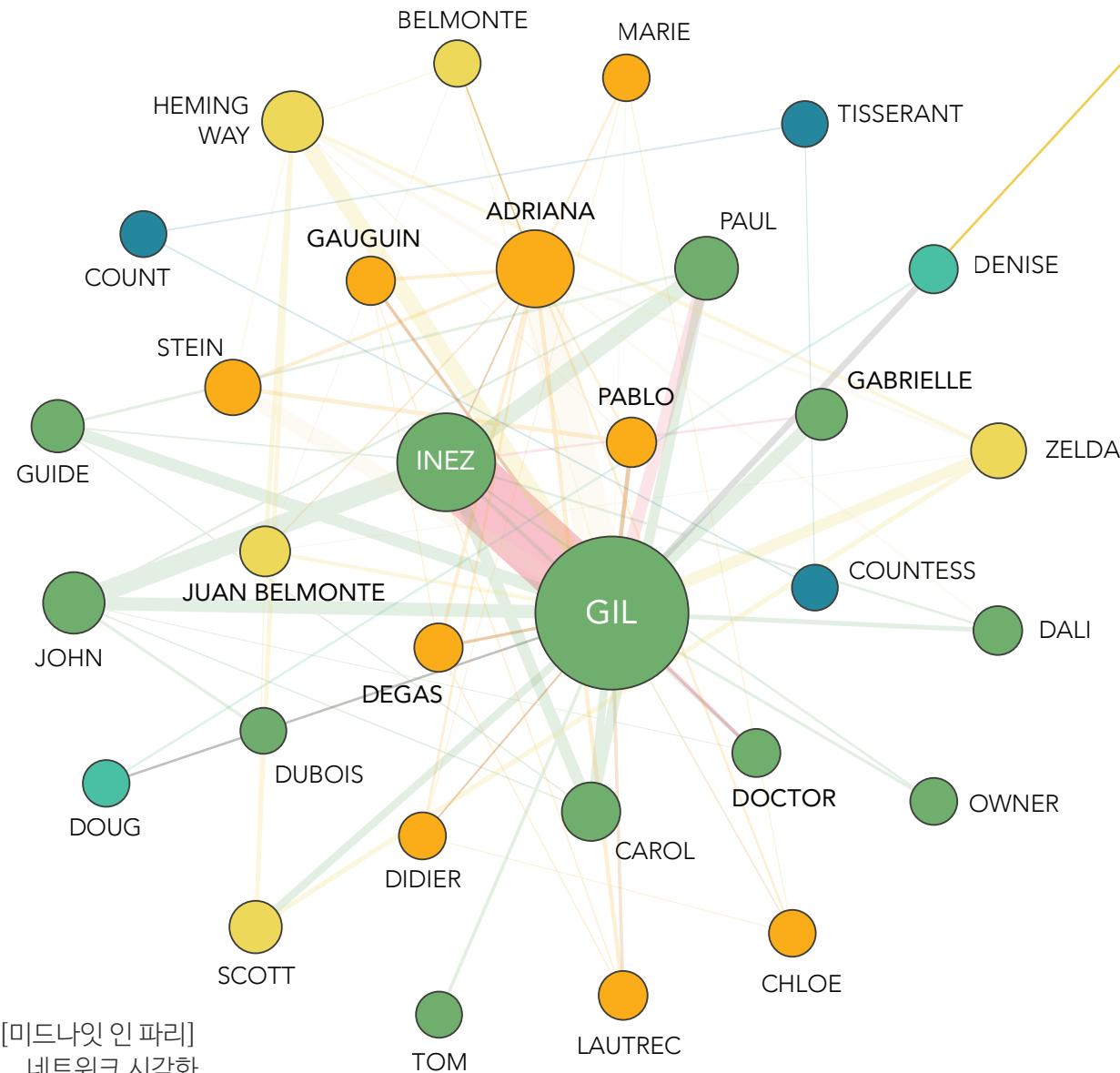


영화 [미드나잇 인 파리] 네트워크 시각화

# BASICS

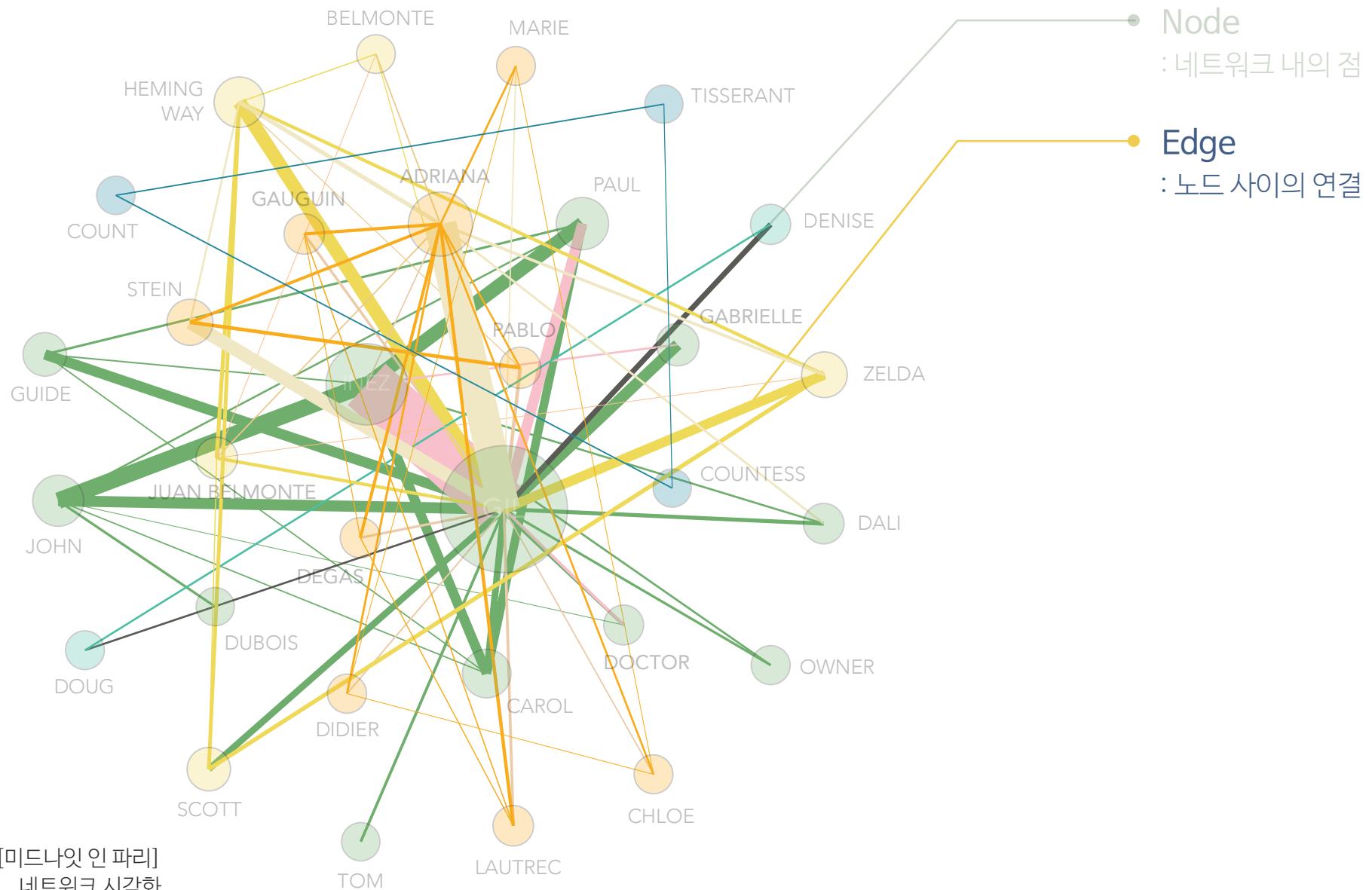
통계량 이해하기

Node  
: 네트워크 내의 점



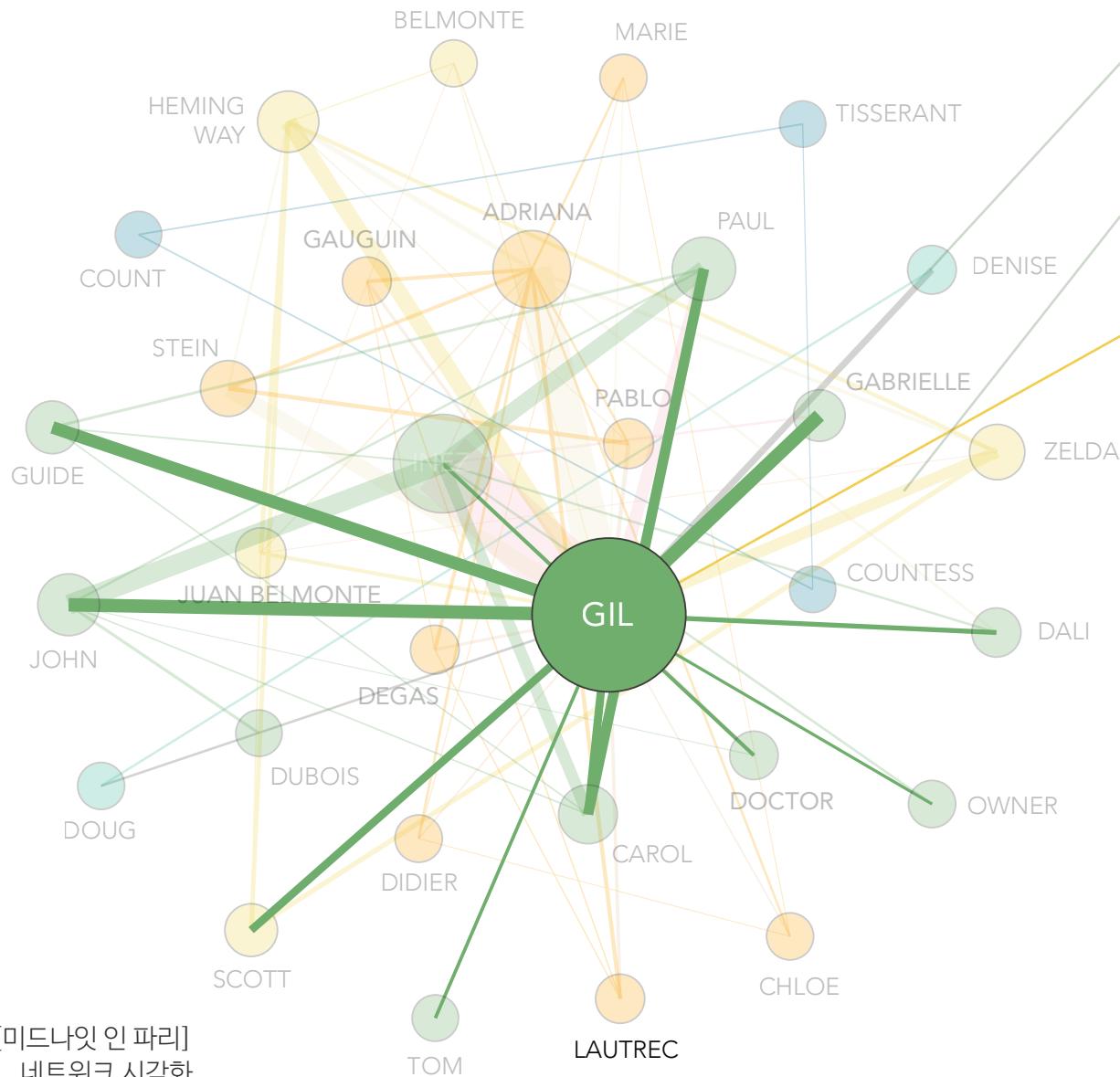
# BASICS

# 통계량 이해하기



# BASICS

## 통계량 이해하기



Node

: 네트워크 내의 점

Edge

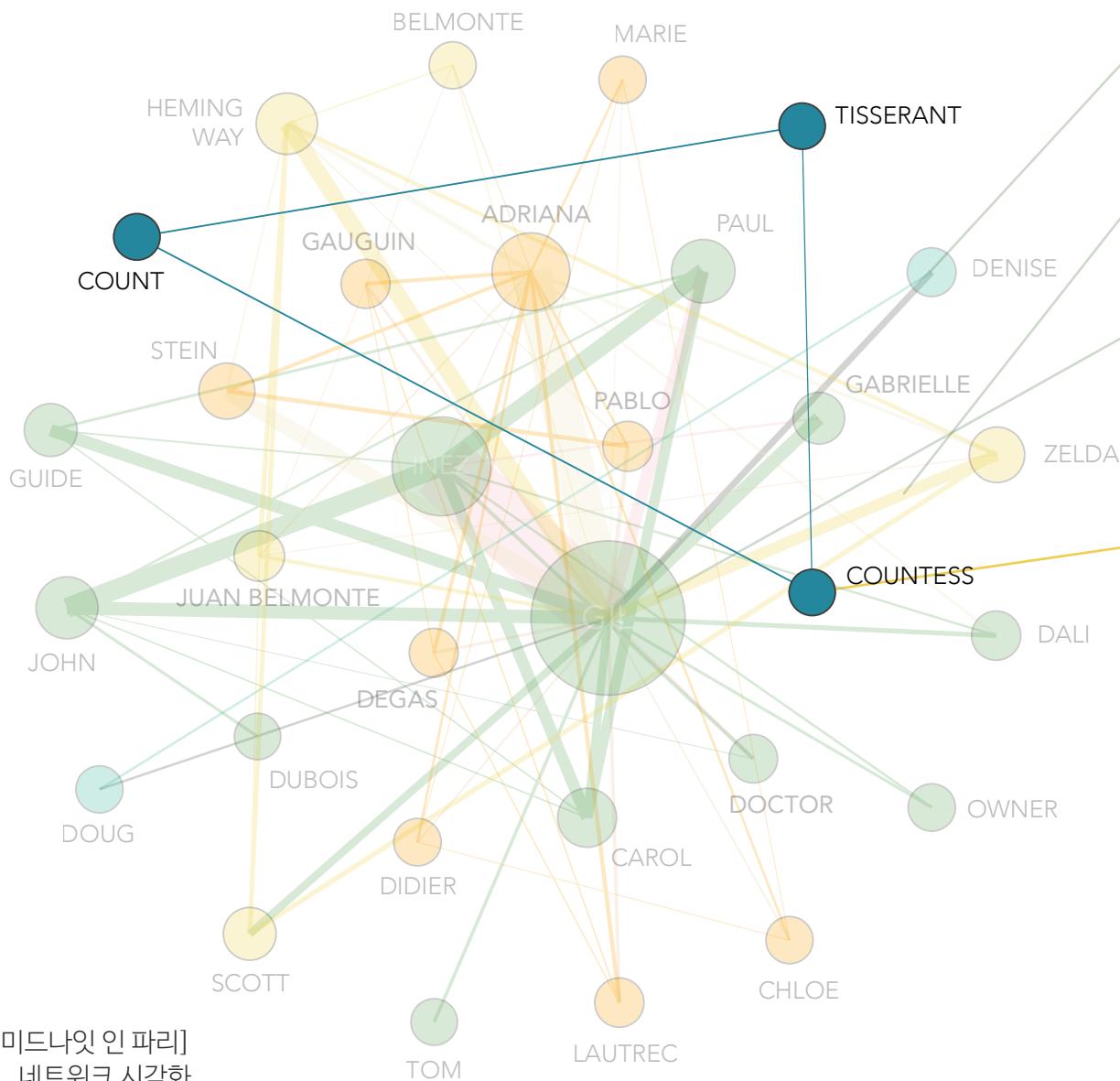
: 노드 사이의 연결

Degree

: 한 Node에 연결된 모든 Edge의 갯수 (Weighted 그래프일 경우에는 모든 Weight의 합)로 중심성을 평가

# BASICS

## 통계량 이해하기



영화 [미드나잇 인 파리]  
네트워크 시각화

### Node

: 네트워크 내의 점

### Edge

: 노드 사이의 연결

### Degree

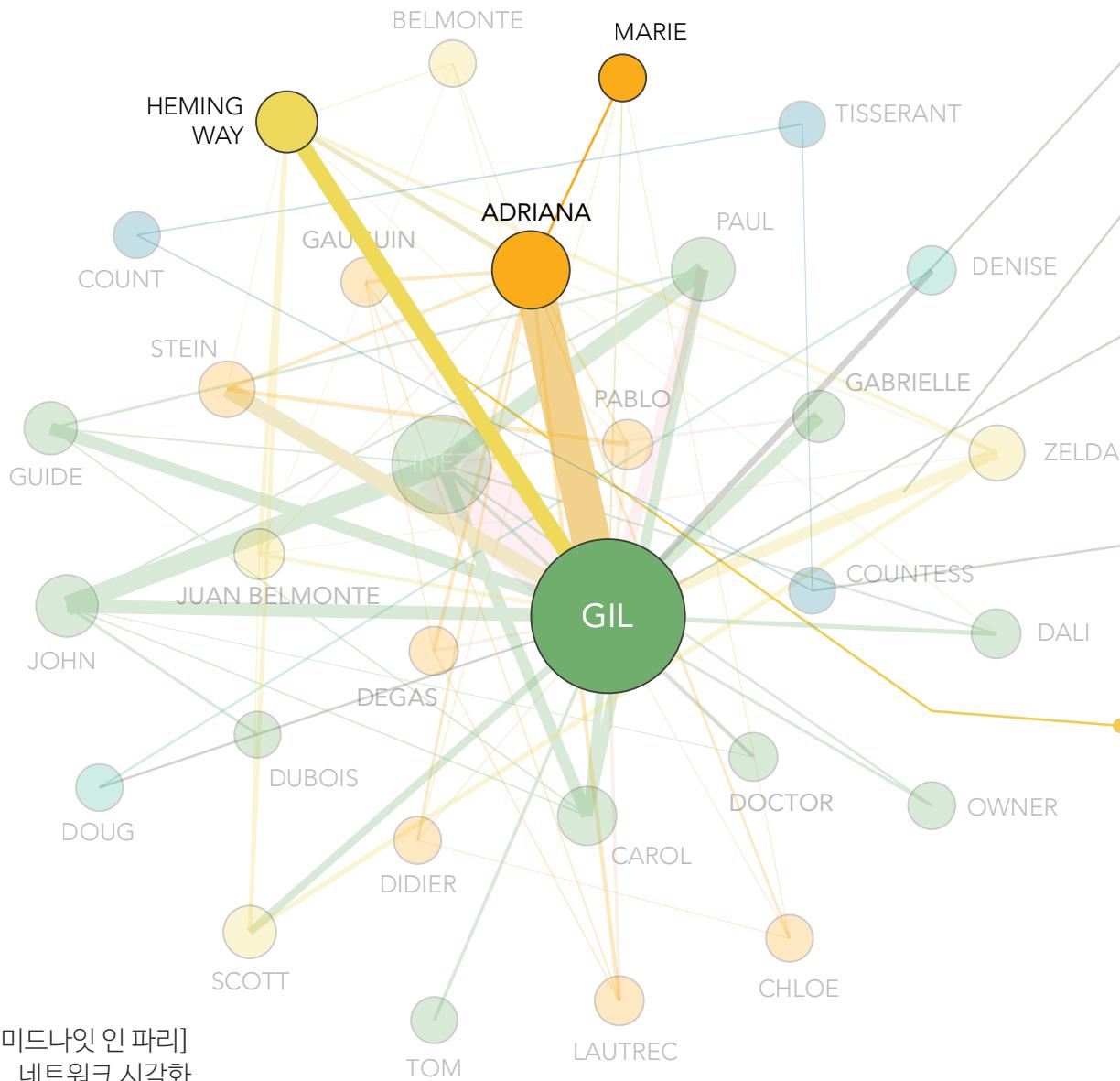
: 한 Node에 연결된 모든 Edge의 갯수 (Weighted 그래프일 경우에는 모든 Weight의 합)로 중심성을 평가

### Betweenness

: A 노드 외 노드에서 다른 노드로 이동 시 얼마나 A를 거쳐가는지 기준으로 A노드 중심성 평가

# BASICS

## 통계량 이해하기



### Node

: 네트워크 내의 점

### Edge

: 노드 사이의 연결

### Degree

: 한 Node에 연결된 모든 Edge의 갯수 (Weighted 그래프일 경우에는 모든 Weight의 합)로 중심성을 평가

### Betweenness

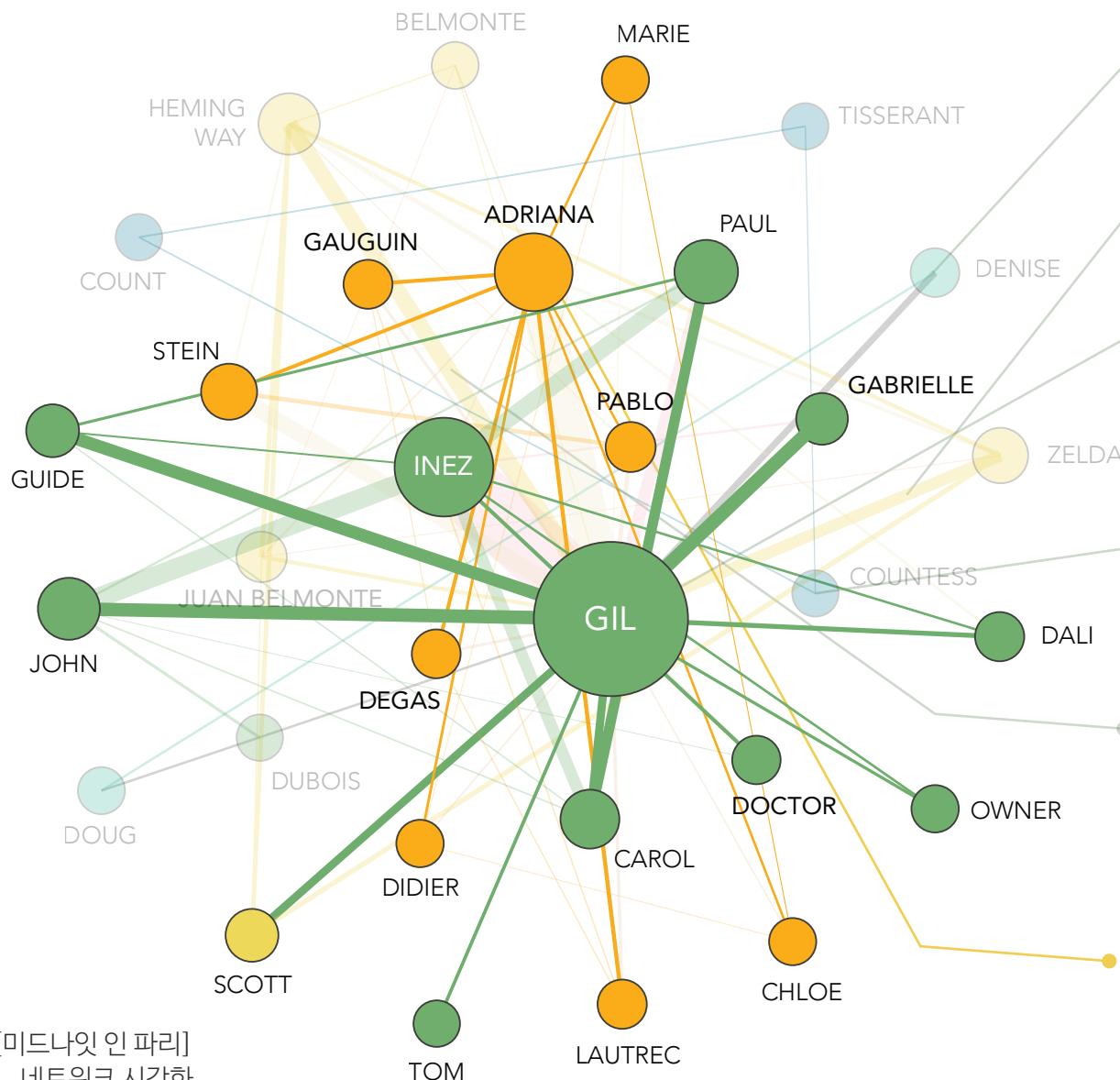
: A 노드 외 노드에서 다른 노드로 이동 시 얼마나 A를 거쳐가는지 기준으로 A노드 중심성 평가

### Closeness

: 중요한 노드일수록 다른 노드까지 도달하는 경로가 짧을 것 (A 노드에서 다른 노드들에 도달하는 시간을 바탕으로 중심성 평가)

# BASICS

## 통계량 이해하기



### Node

: 네트워크 내의 점

### Edge

: 노드 사이의 연결

### Degree

: 한 Node에 연결된 모든 Edge의 갯수 (Weighted 그래프일 경우에는 모든 Weight의 합)로 중심성을 평가

### Betweenness

: A 노드 외 노드에서 다른 노드로 이동 시 얼마나 A를 거쳐가는지 기준으로 A노드 중심성 평가

### Closeness

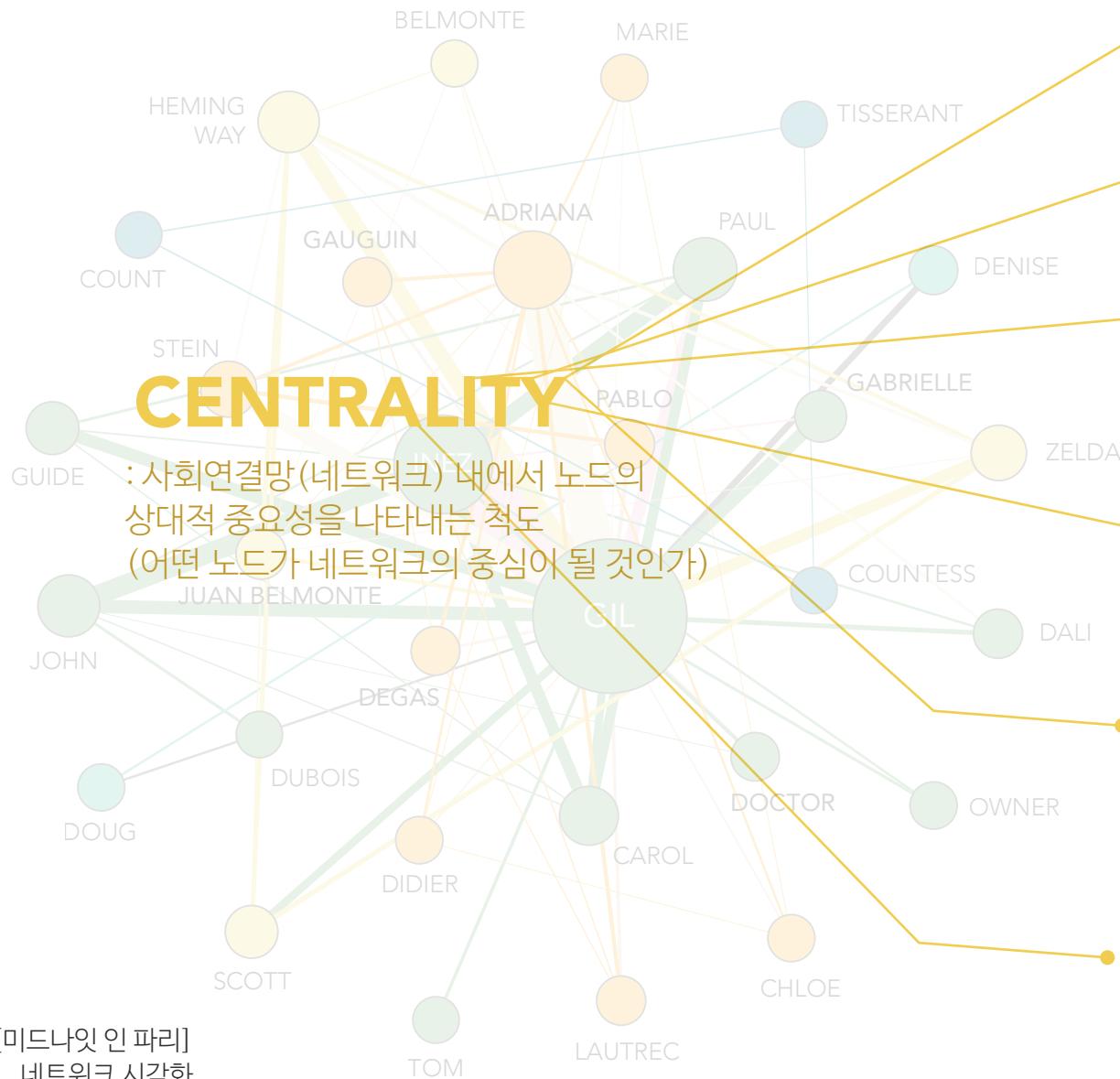
: 중요한 노드일수록 다른 노드까지 도달하는 경로가 짧을 것 (A 노드에서 다른 노드들에 도달하는 시간을 바탕으로 중심성 평가)

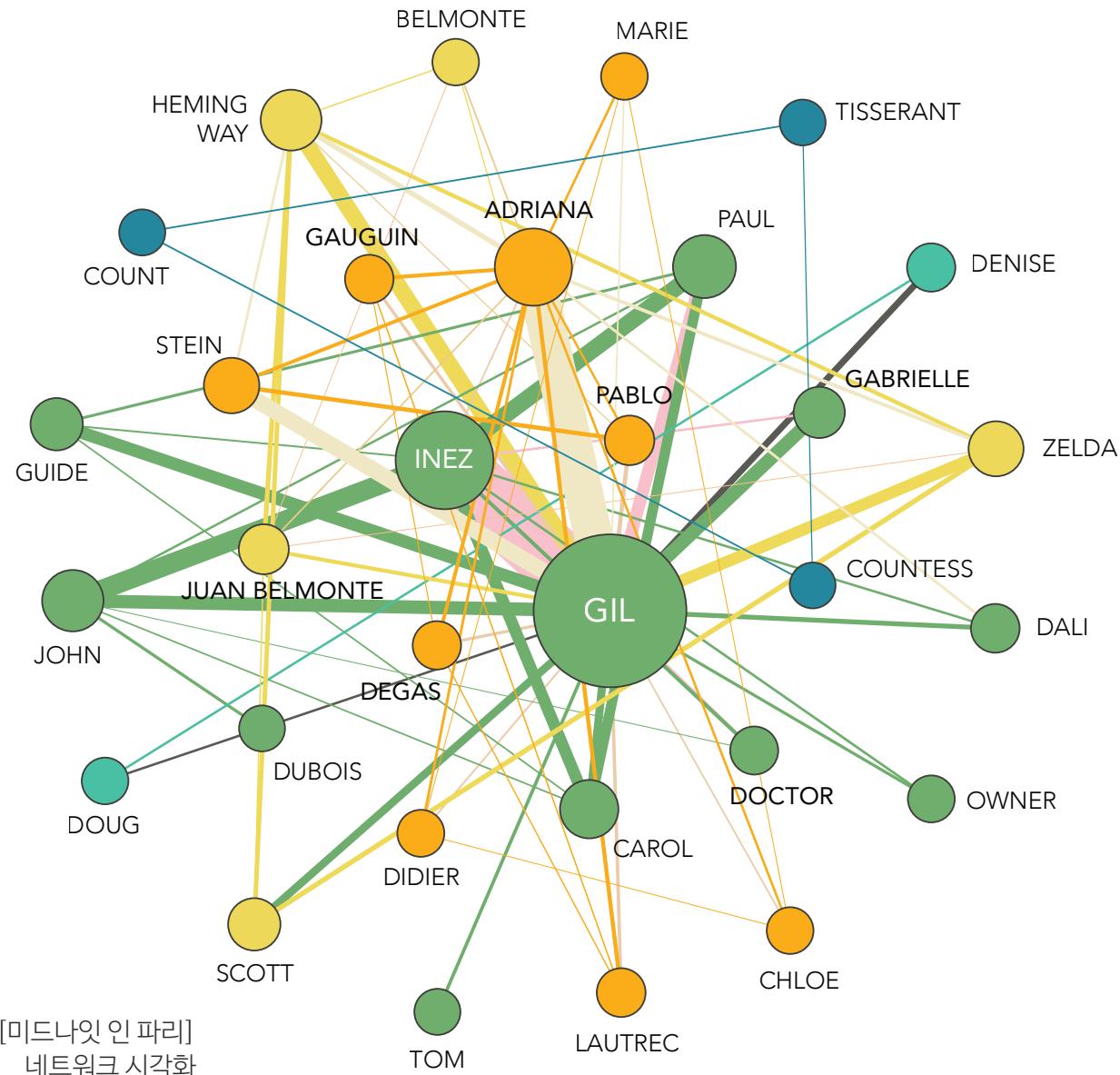
### Eigenvector

: 연결된 노드가 많다고 무조건 중요한 노드가 되는 것은 아님 (중요한 노드와의 연결 여부로 중심성 평가)

# BASICS

## 통계량 이해하기

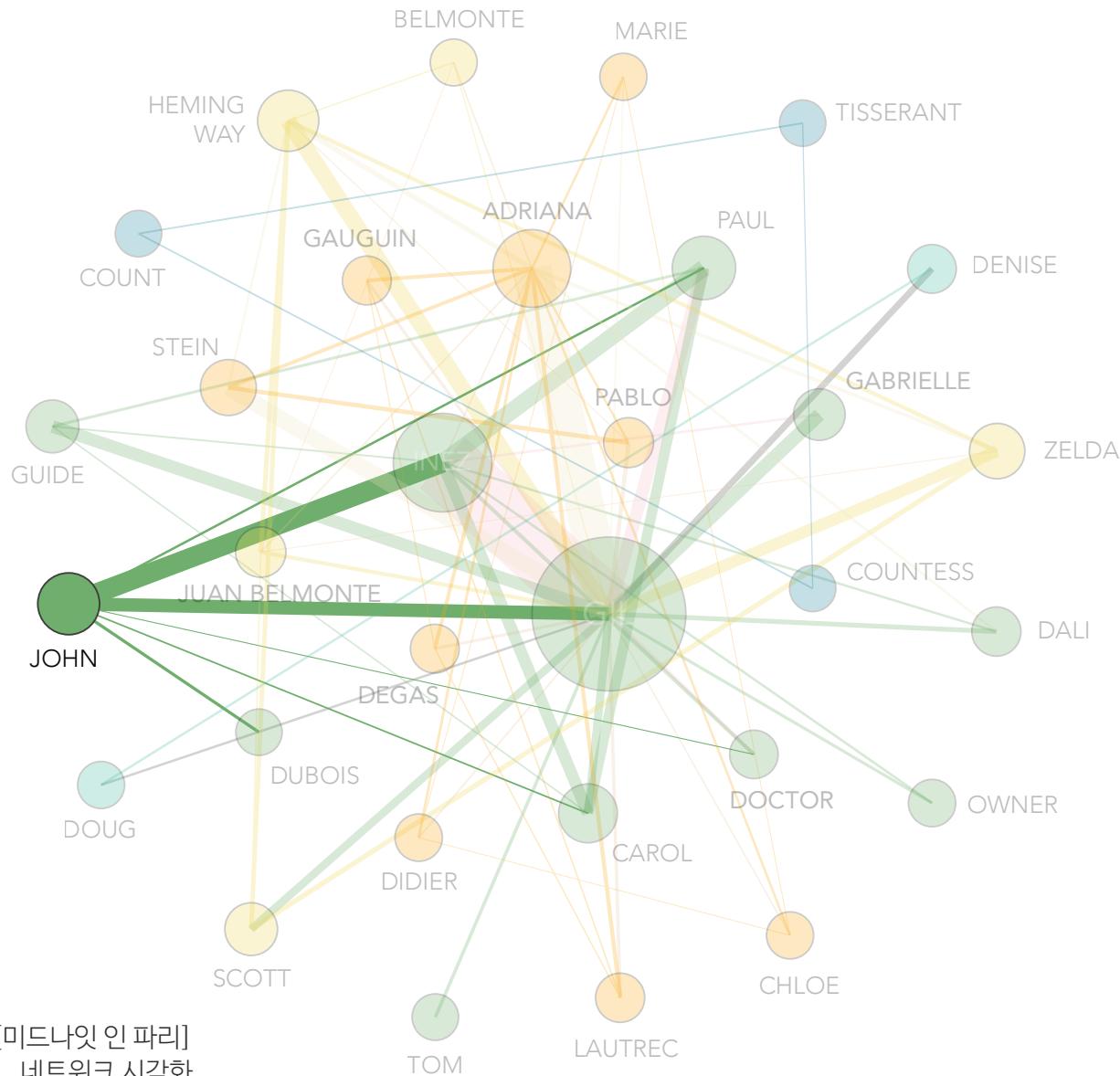




## 영화 [미드나잇 인 파리] 네트워크 시각화

# BASICS

통계량 이해하기



## Edge 생성기준

: 두 인물이 동일 장면 (scene)에서 등장할 경우  
(두 인물의 co-occurrence)

## Edge-Weight

1. 두 인물의 동시 등장 횟수  
(동시 등장한 scene과 정비례 관계)
2. 각 장면에 등장한 인물의 수  
(2인 대화와 같이 긴밀한 대화일수록 weight가 커지고, 다자 대화일수록 작아진다)

## BASICS

통계량 이해하기

### Edge-Weight

: 다자 대화에서 각 인물 관계에 주어지는 weight는 어떻게 결정 될까?



CHARACTER A



CHARACTER B



CHARACTER C



CHARACTER D

이야기에서 등장인물 A, B, C, D가 출연한다고 했을 때,  
각각 A, B, C, D끼리는 서로 대화를 나눴지만, 발화의 횟수는 다르다.

## BASICS

통계량 이해하기

CHARACTER A



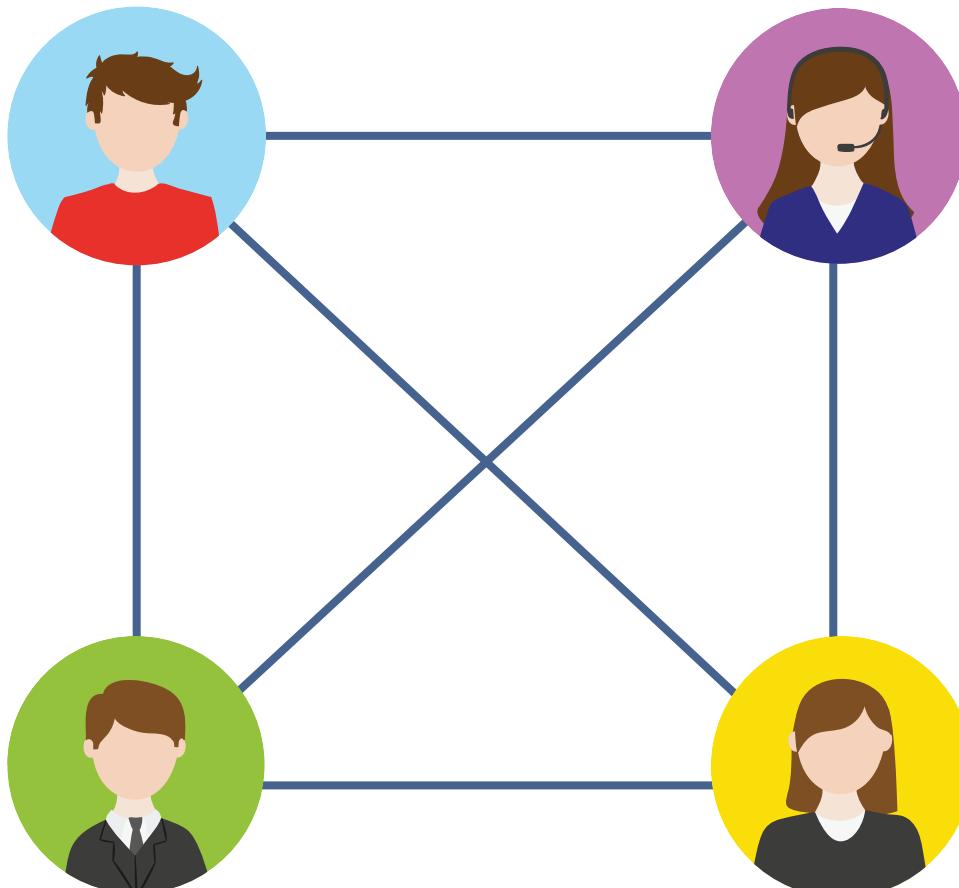
CHARACTER B



CHARACTER C

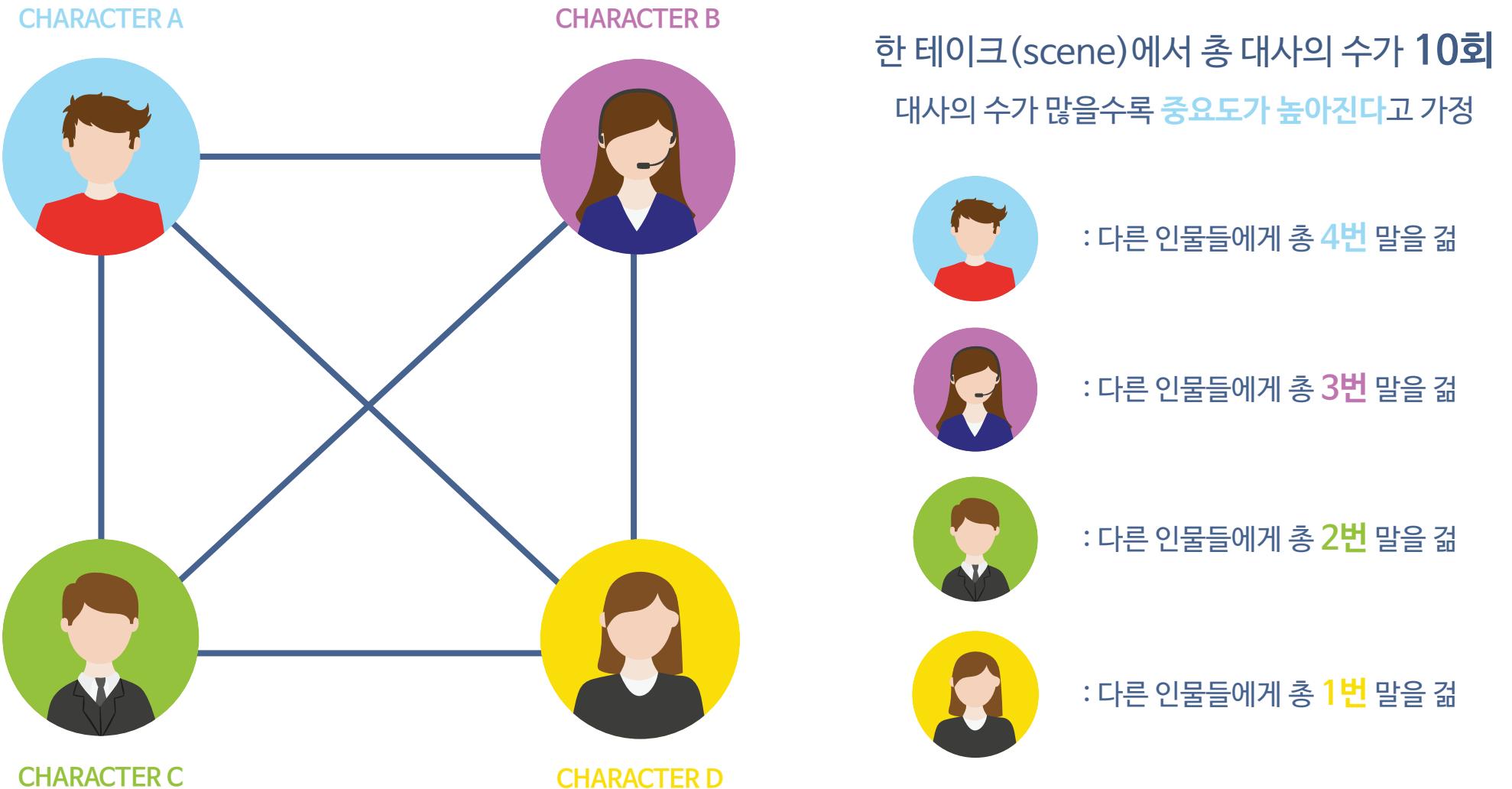


CHARACTER D



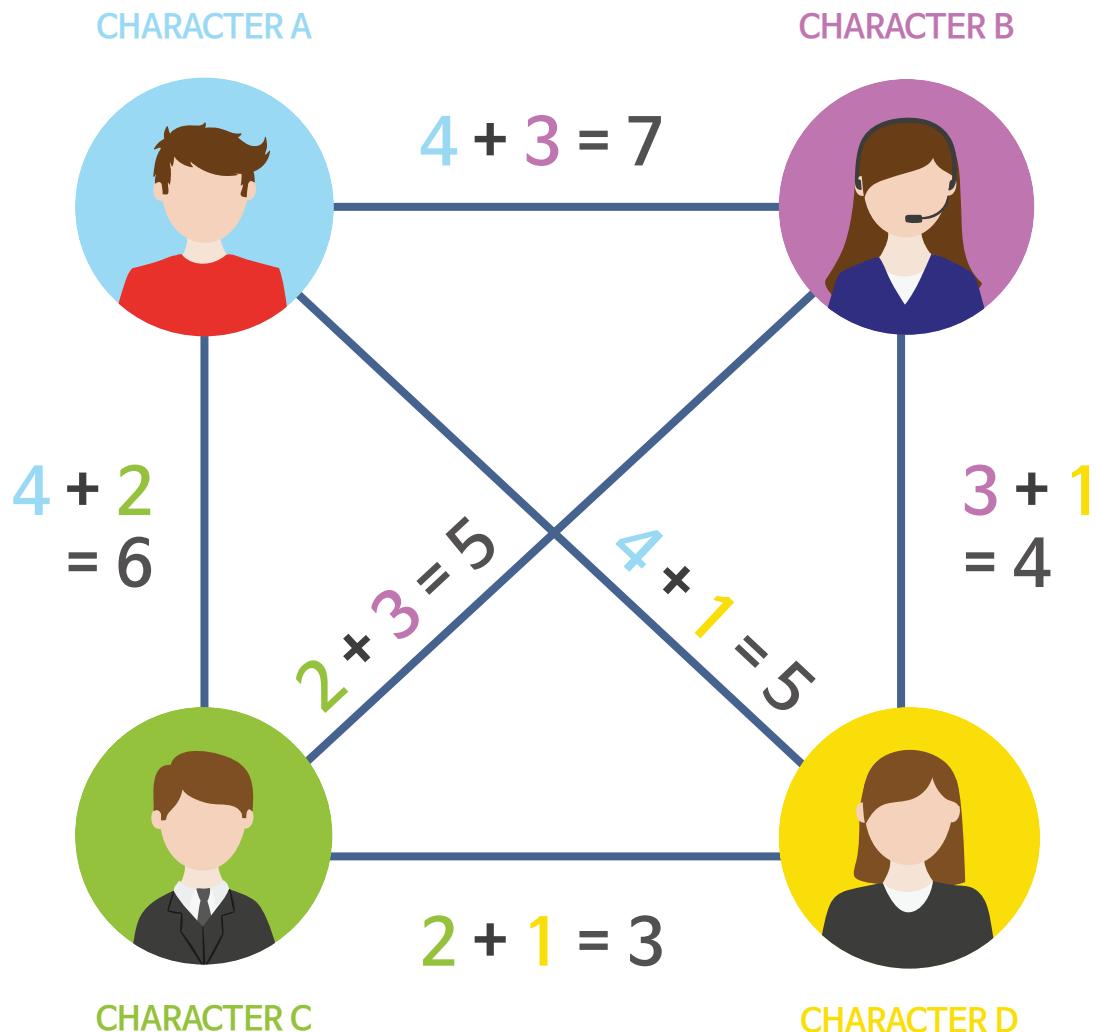
## BASICS

### 통계량 이해하기



## BASICS

통계량 이해하기



캐릭터 A 와 캐릭터 B 사이의 edge-weight =

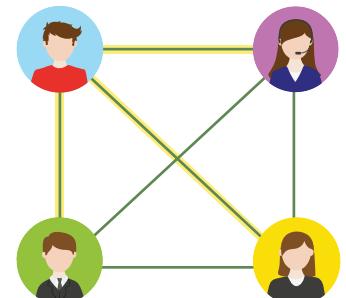
$$\frac{4 + 3 = 7}{\text{캐릭터 A 대사 수} + \text{캐릭터 B 대사 수}} \times \sqrt{10}$$

총 대사의 수

$$10 \times (4 - 1)$$

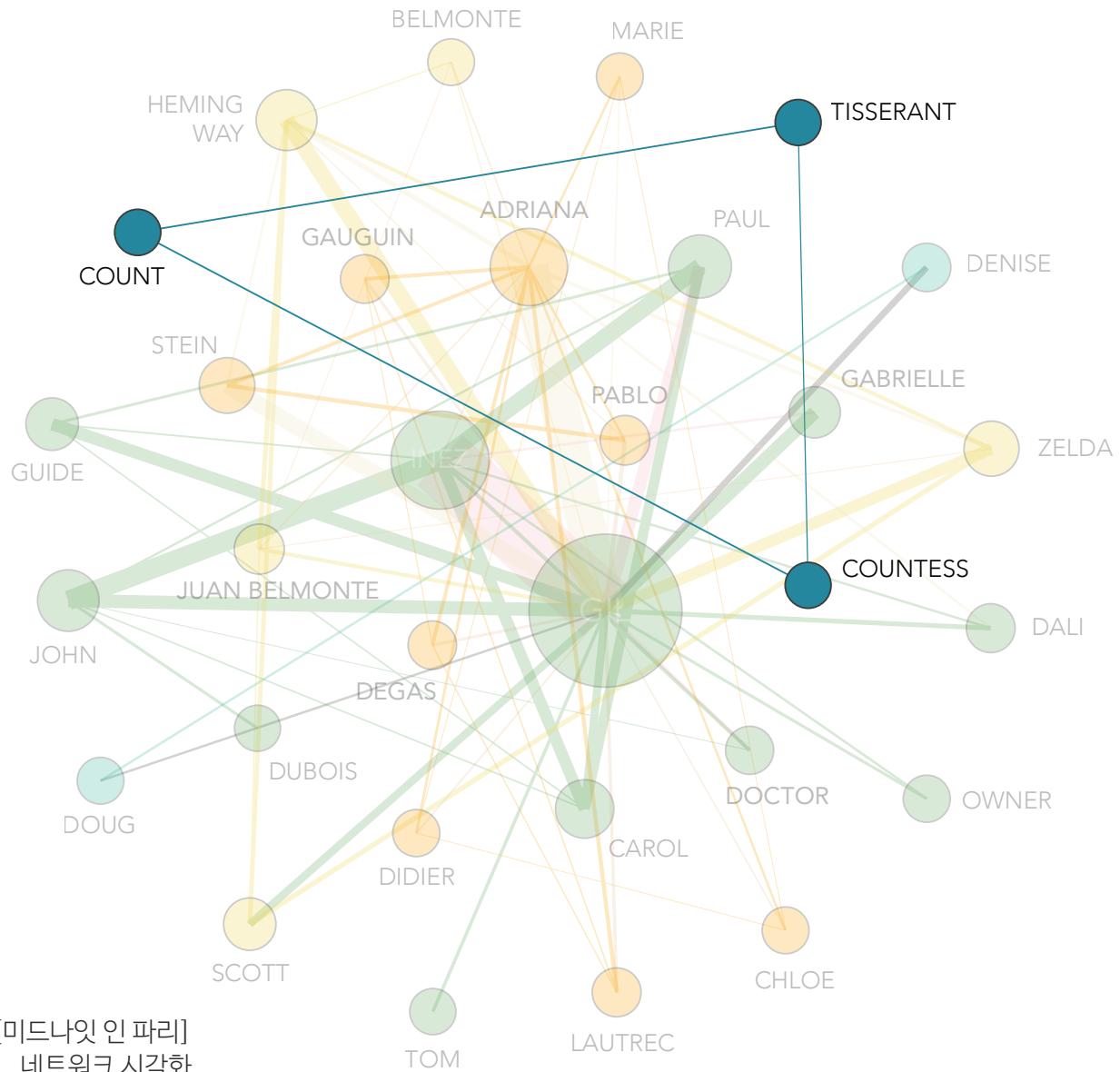
총 대사의 수  $\times$  (캐릭터의 수 - 1)

: 한 캐릭터 당 연결되어 있는 인물의 수는  
자신을 제외한 3명



# BASICS

통계량 이해하기



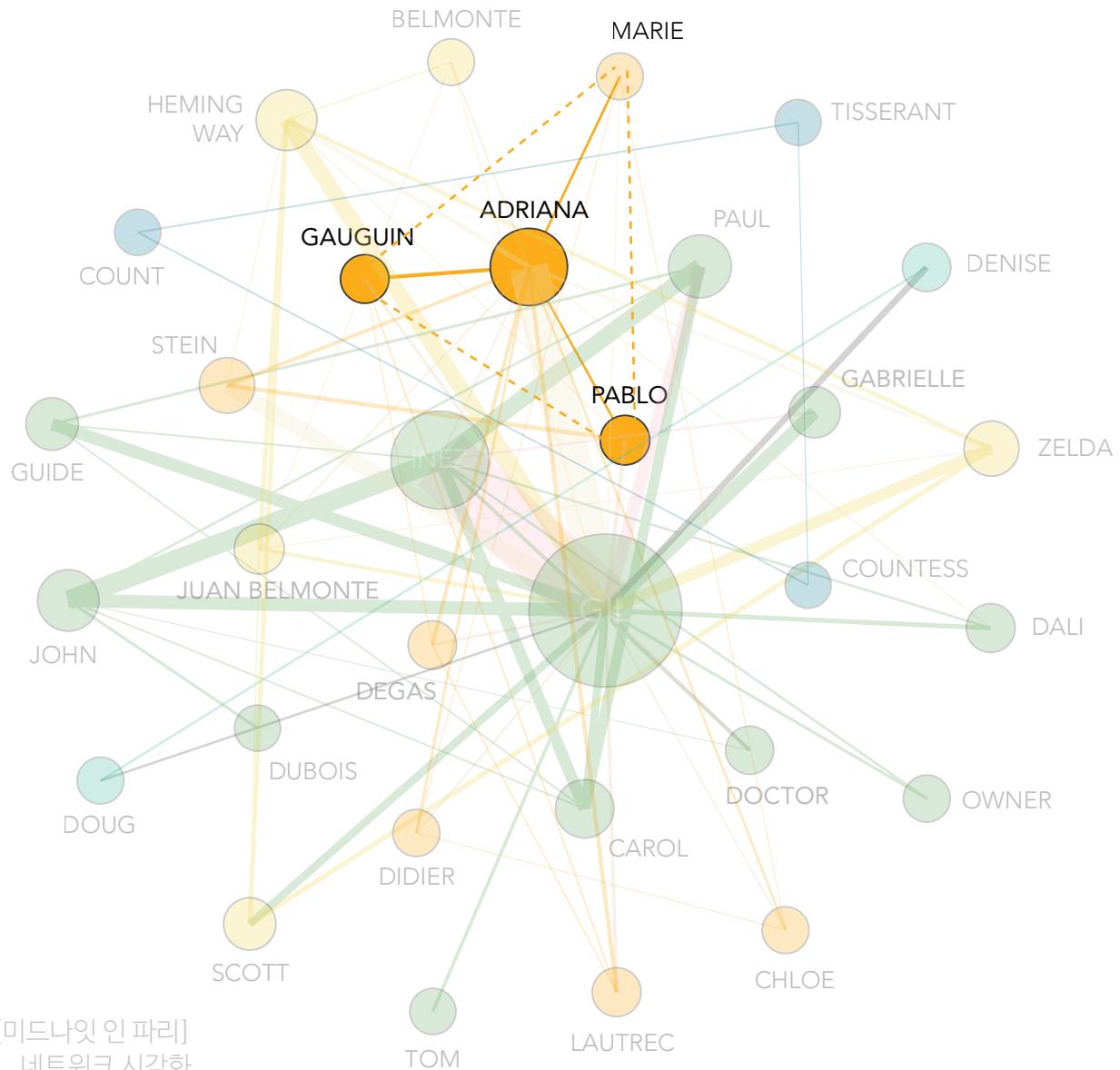
## 밀도 (Density)

$$\frac{\text{실제 엣지 수 (3)}}{\text{가능한 엣지 수 (3)}}$$

= 1

# BASICS

통계량 이해하기



밀도 (Density)

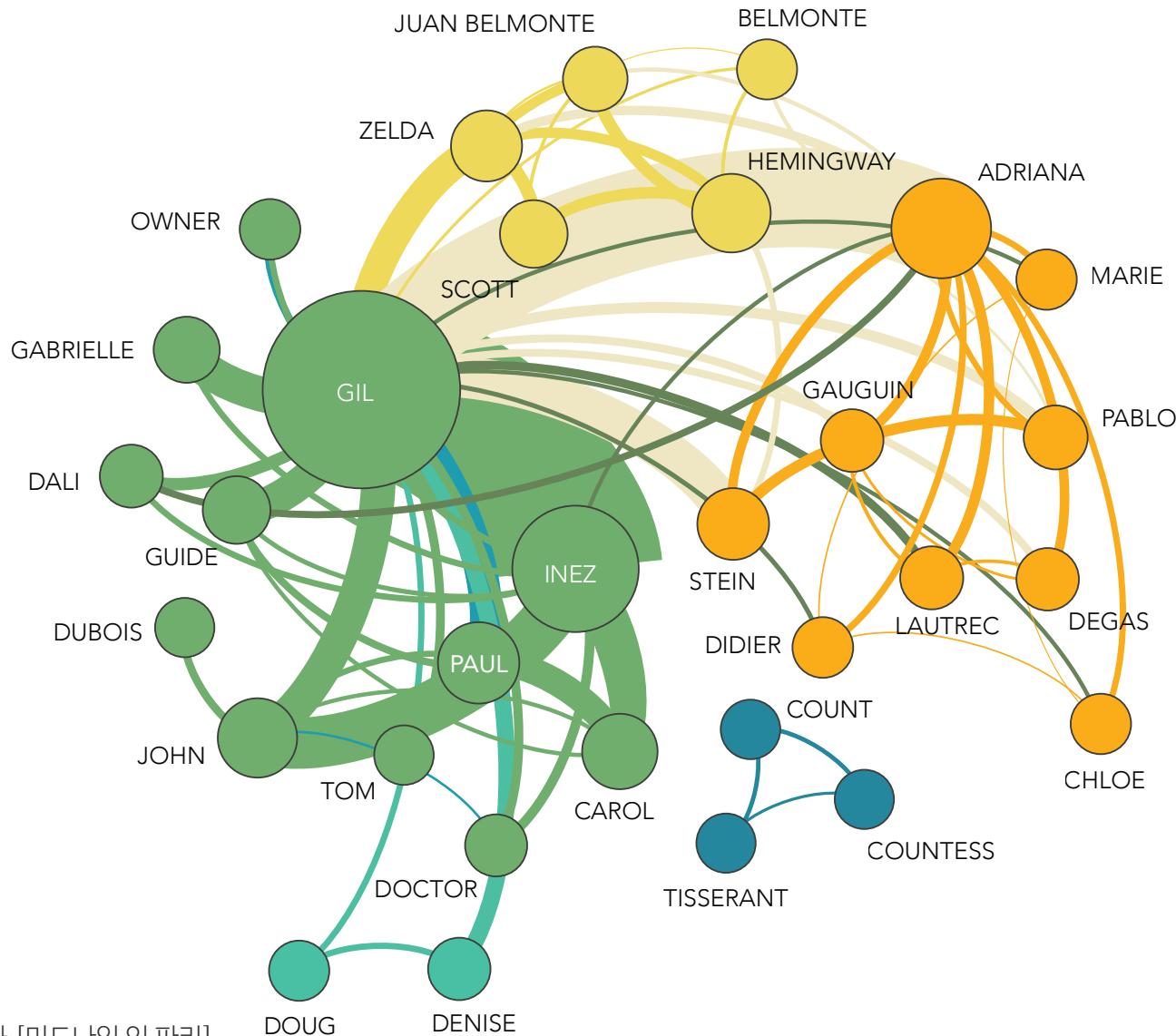
실제 엣지 수 (3)

가능한 엣지 수 (6)

$$= 0.5$$

## BASICS

### 통계량 이해하기



영화 [미드나잇 인 파리]  
네트워크 시각화

## 모듈성 (Modularity)

: 한 그래프에서 노드끼리  
서로 그룹화되어 있을 확률

: 좌측 그래프에서는 모듈성을  
각각 색깔로 구분

: 전체 네트워크 안에서 커뮤니티를  
가정하였을 때, 커뮤니티 내부에 펼쳐져  
있는 연결들이 무작위적인 연결들과  
비교했을 때 얼마나 더 많은지를 정량화

## BASICS

통계량 이해하기

NUMBER OF EACH GROUP

VARIANCE DEGREE

MAXIMUM PAGE RANK

DENSITY

MEAN CHART FREQUENCY

MEAN HARMONIC CLOSENESS

NUMBER OF MODULARITY

SCENE LENGTH

VARIANCE CHART FREQUENCY

VARIANCE BETWEENNESS

MEAN DEGREE

NUMBER OF NODES

VARIANCE DIAMETER

MAXIMUM BETWEENNESS

MAX WEIGHTED DEGREE

2ND MAX CHART FREQUENCY

NUMBER OF LINKS

MEAN SCENE LENGTH

MAX SCENE LENGTH

MAX DIAMETER

VARIANCE WEIGHTED DEGREE

VARIANCE SCENE LENGTH

MEAN DIAMETER

MAX CHART FREQUENCY

2ND MAX WEIGHTED DEGREE

MAX DEGREE

2ND MAXIMUM DEGREE

MEAN CLOSENESS

SUCCESS

## RESULTS

통계량 분석하기

### 정규성 검정 (Shapiro Test)

정규성 만족 변수 20개

등분산 검정 (Bartlett Test)

Two Sample T-test ( $\alpha = 0.2$ )

정규성 불만족 변수 20개

등분산 검정 (Levene Test)

Wilcoxon Signed-Rank Test ( $\alpha = 0.2$ )

SECOND MAX CHARACTER FREQUENCY

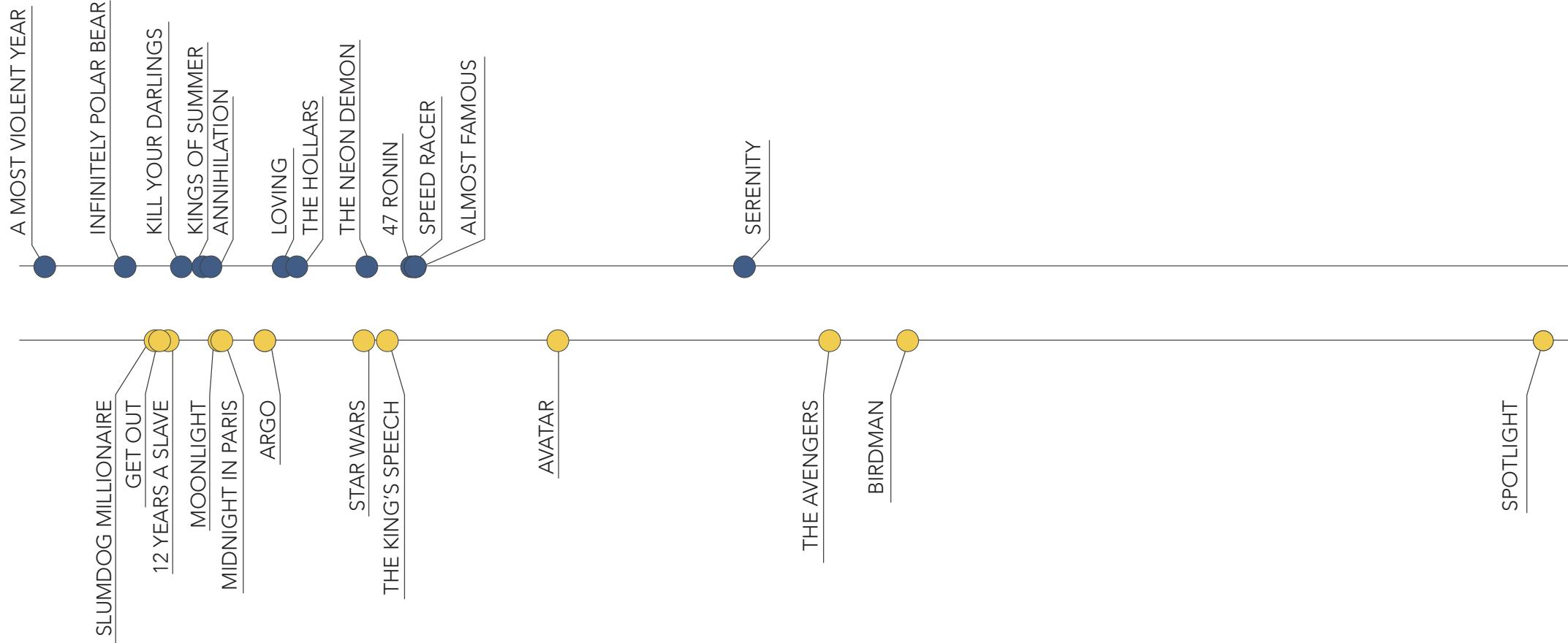
VARIANCE CHARACTER FREQUENCY

MEAN DEGREE

# RESULTS

분석결과

차이가 가장 큰 값: mean degree



# RESULTS

분석결과

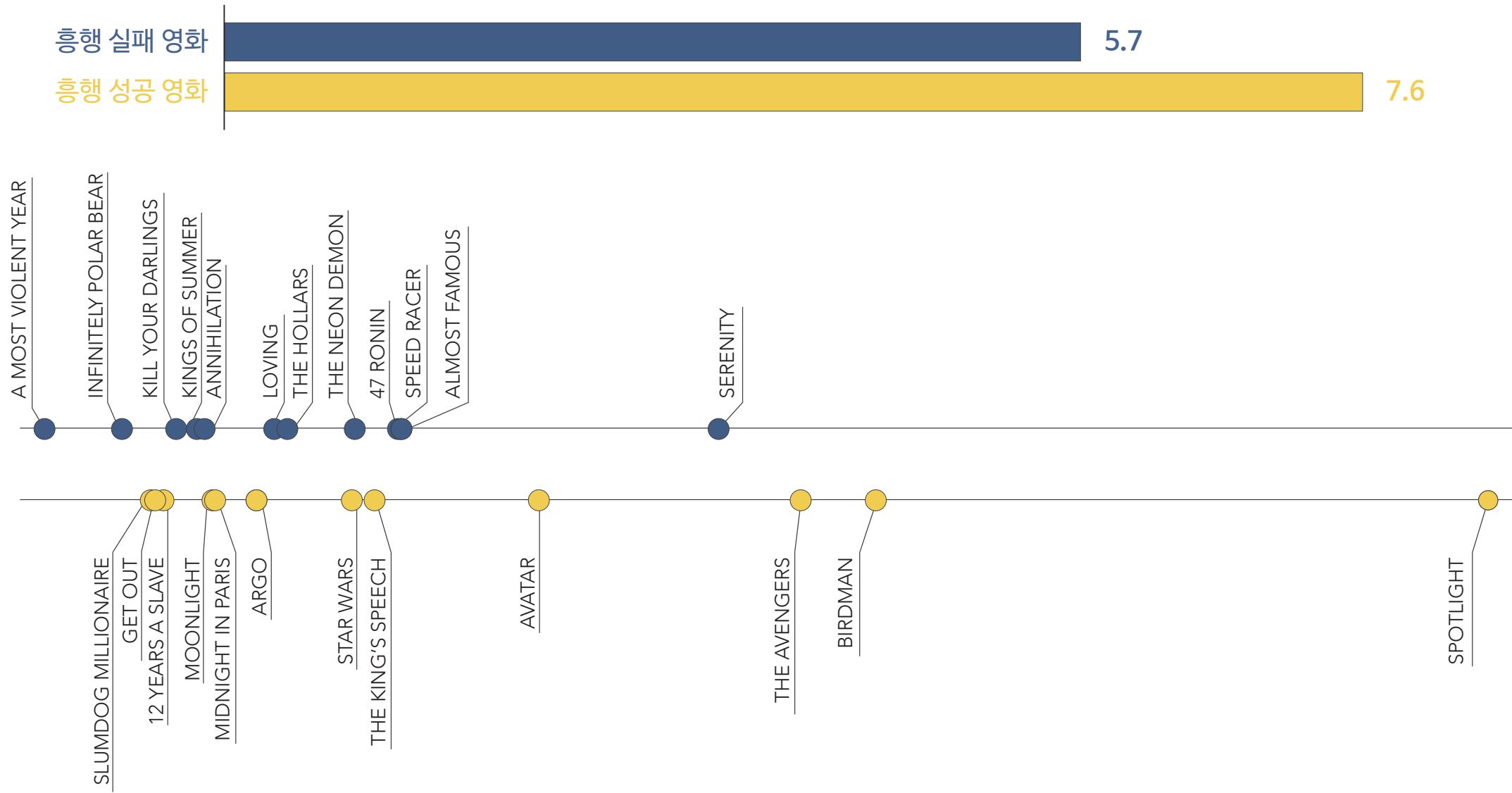
차이가 가장 큰 값: mean degree



# RESULTS

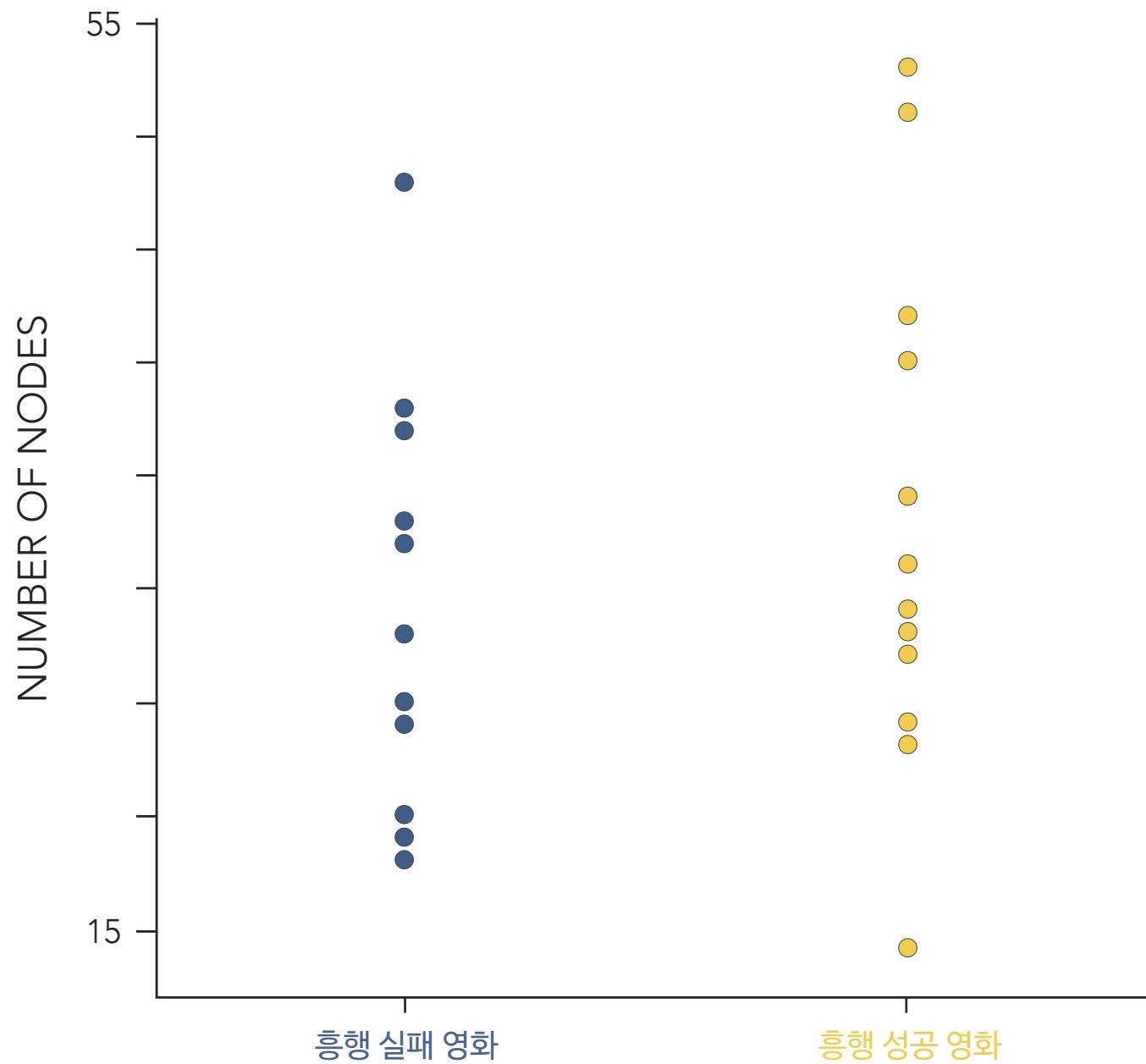
분석결과

차이가 가장 큰 값: mean degree



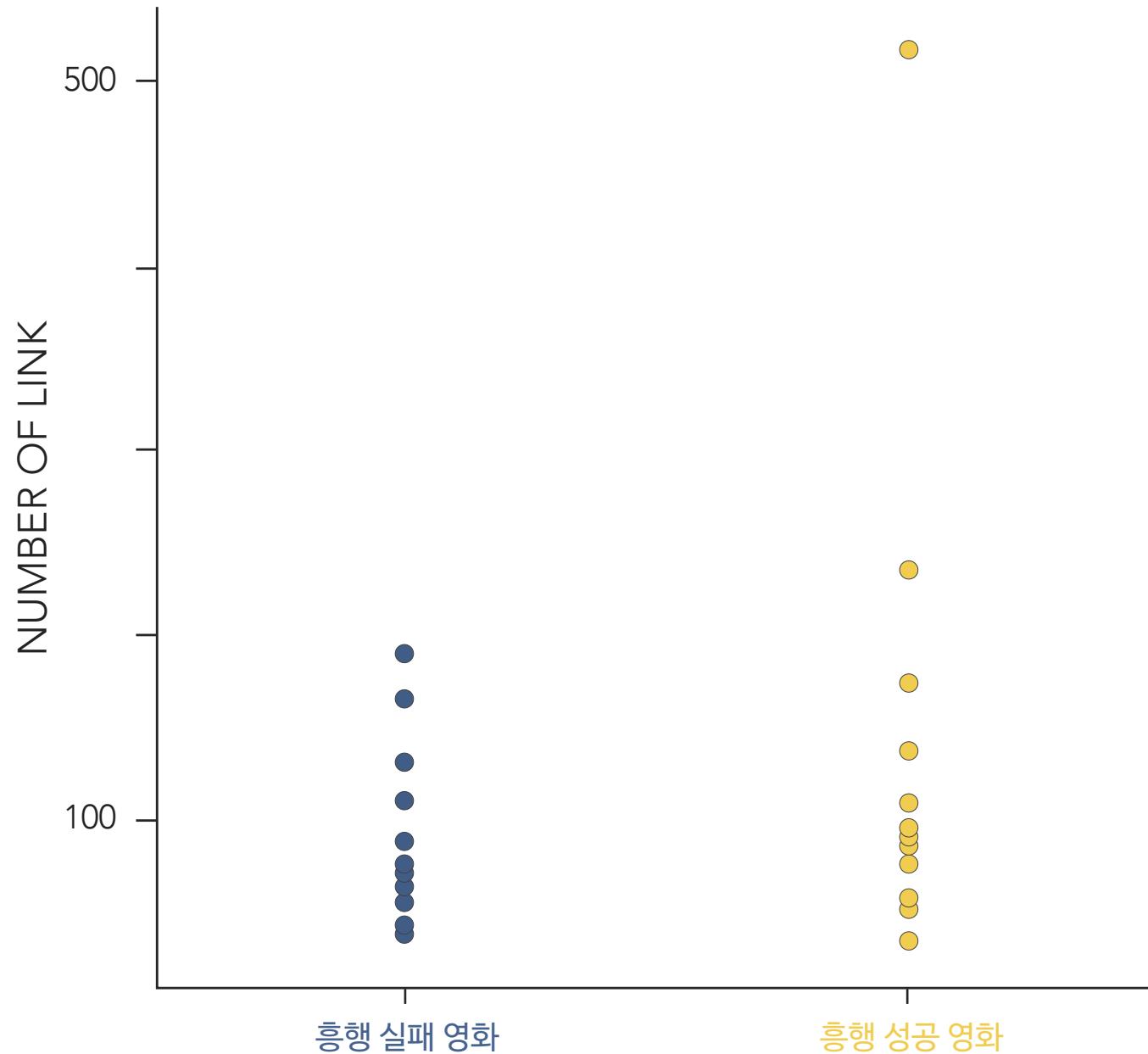
# RESULTS

통계량 시각화 결과



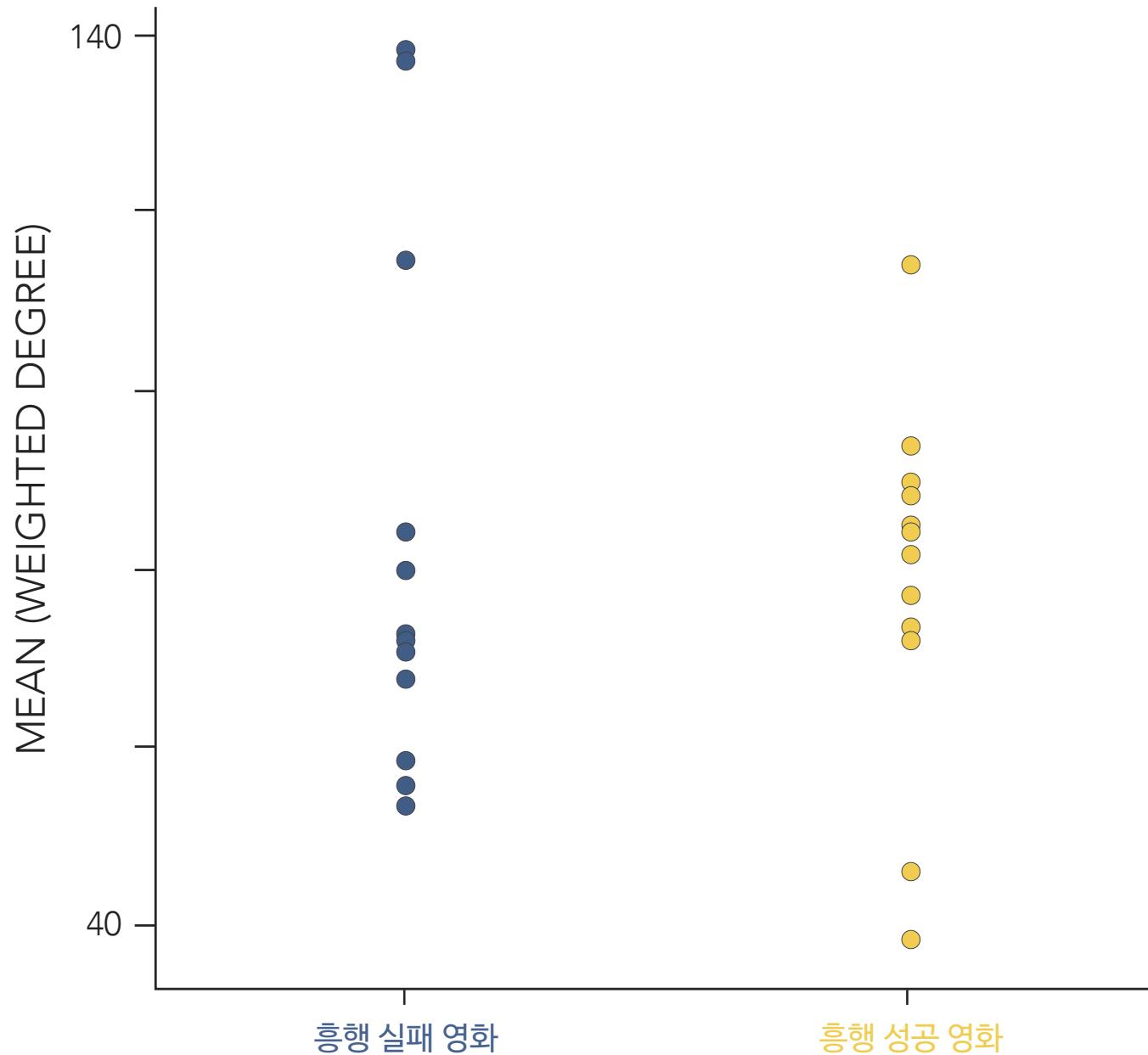
## RESULTS

통계량 시각화 결과



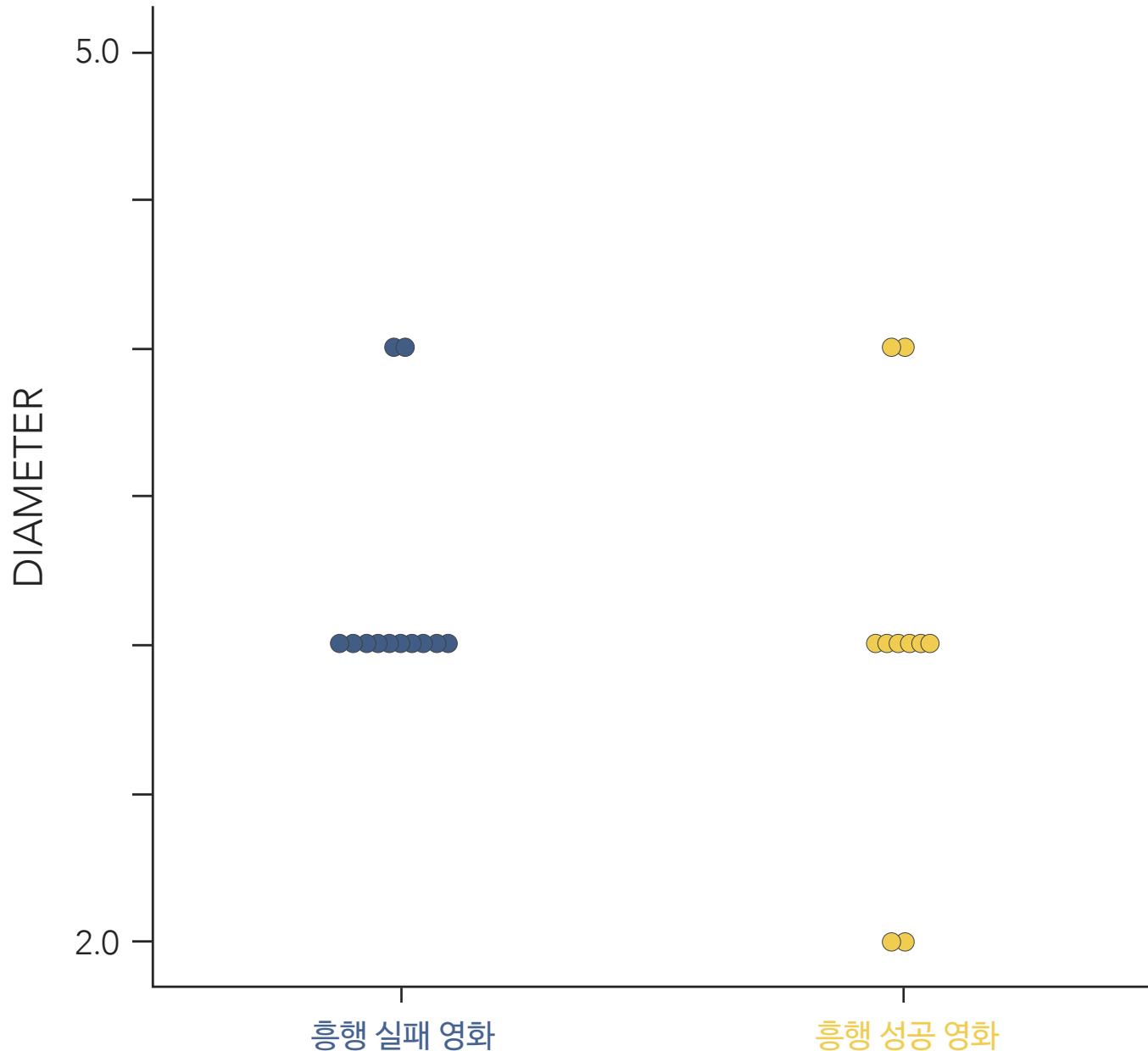
## RESULTS

통계량 시각화 결과



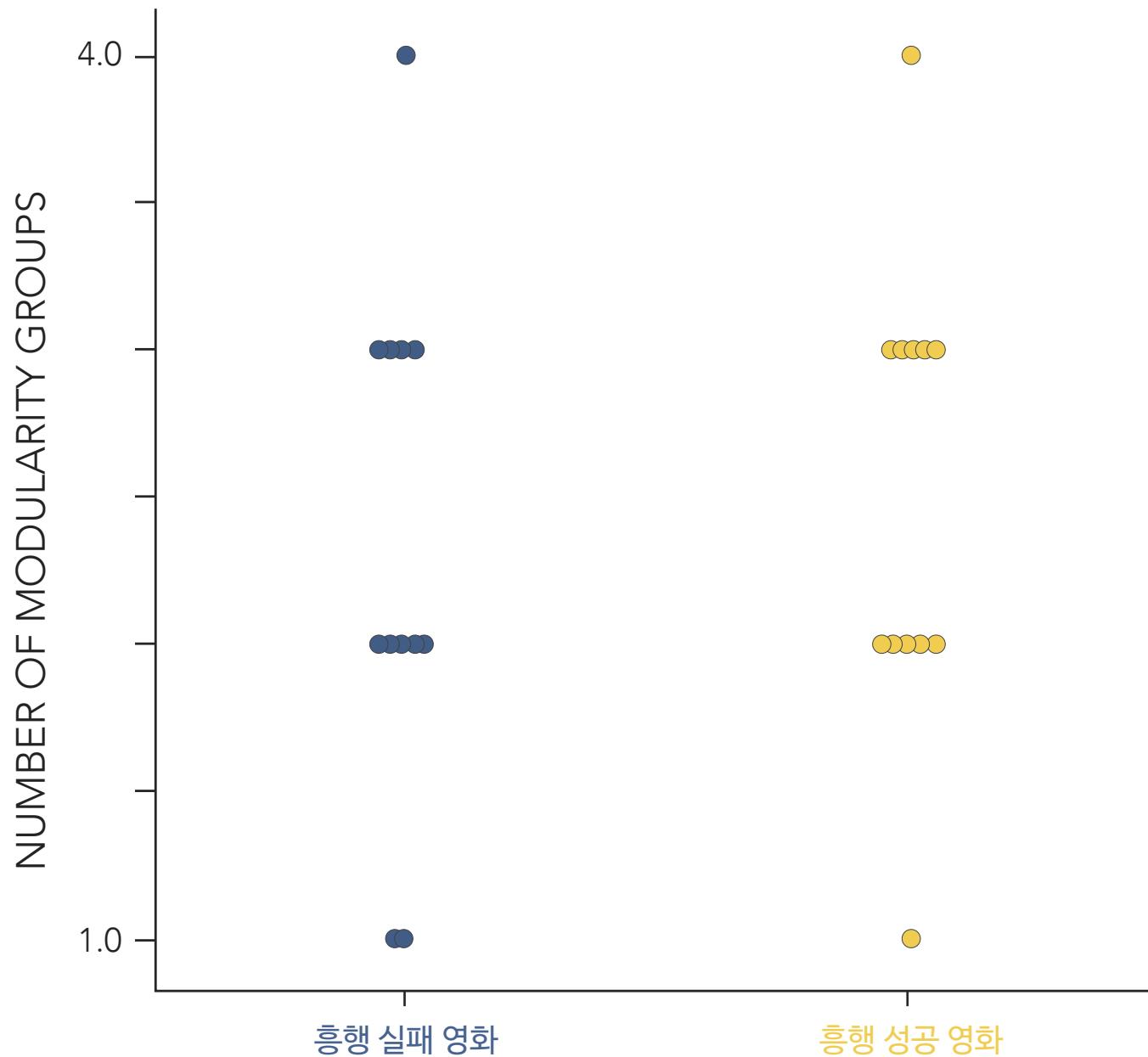
## RESULTS

통계량 시각화 결과



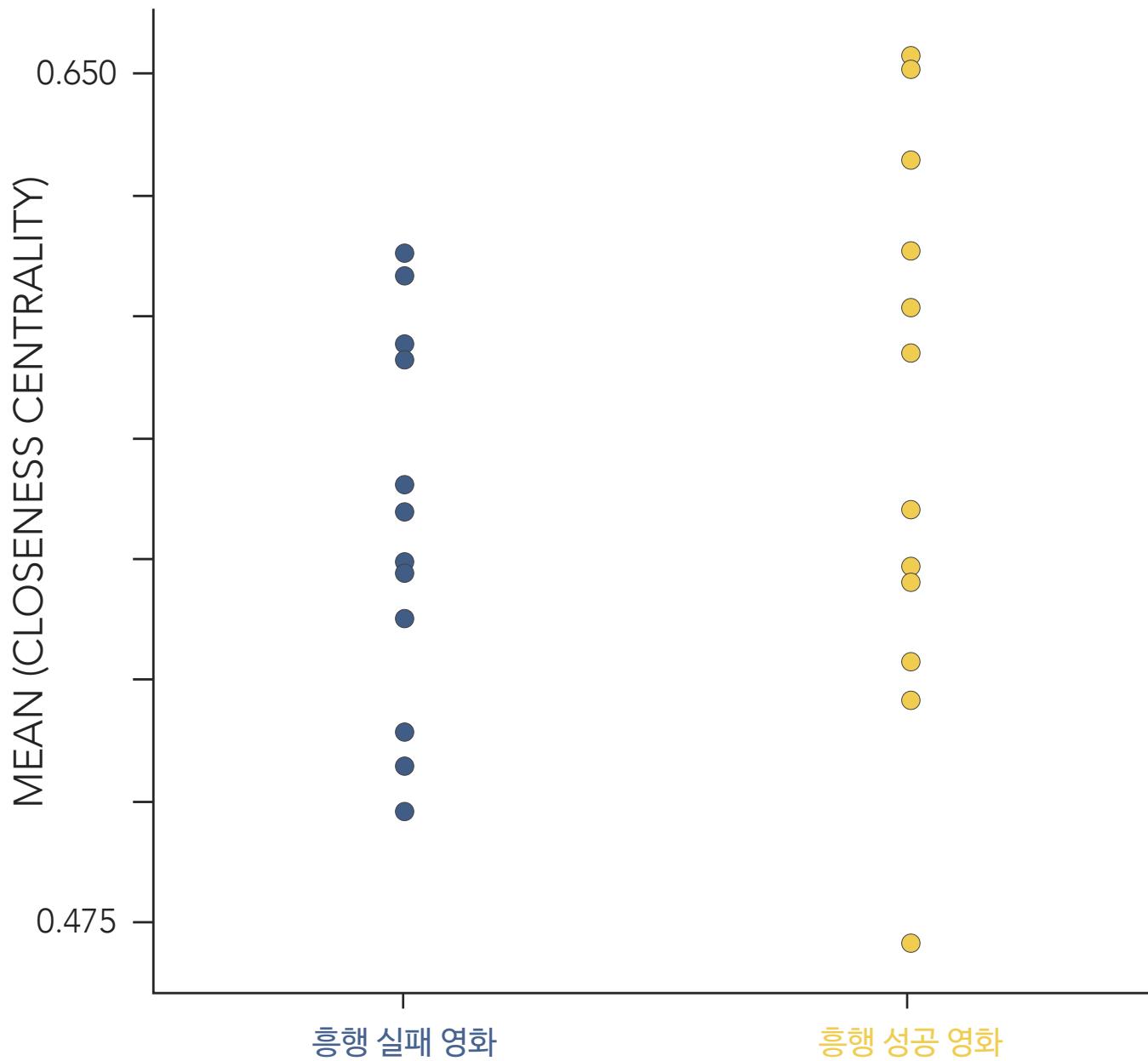
## RESULTS

통계량 시각화 결과



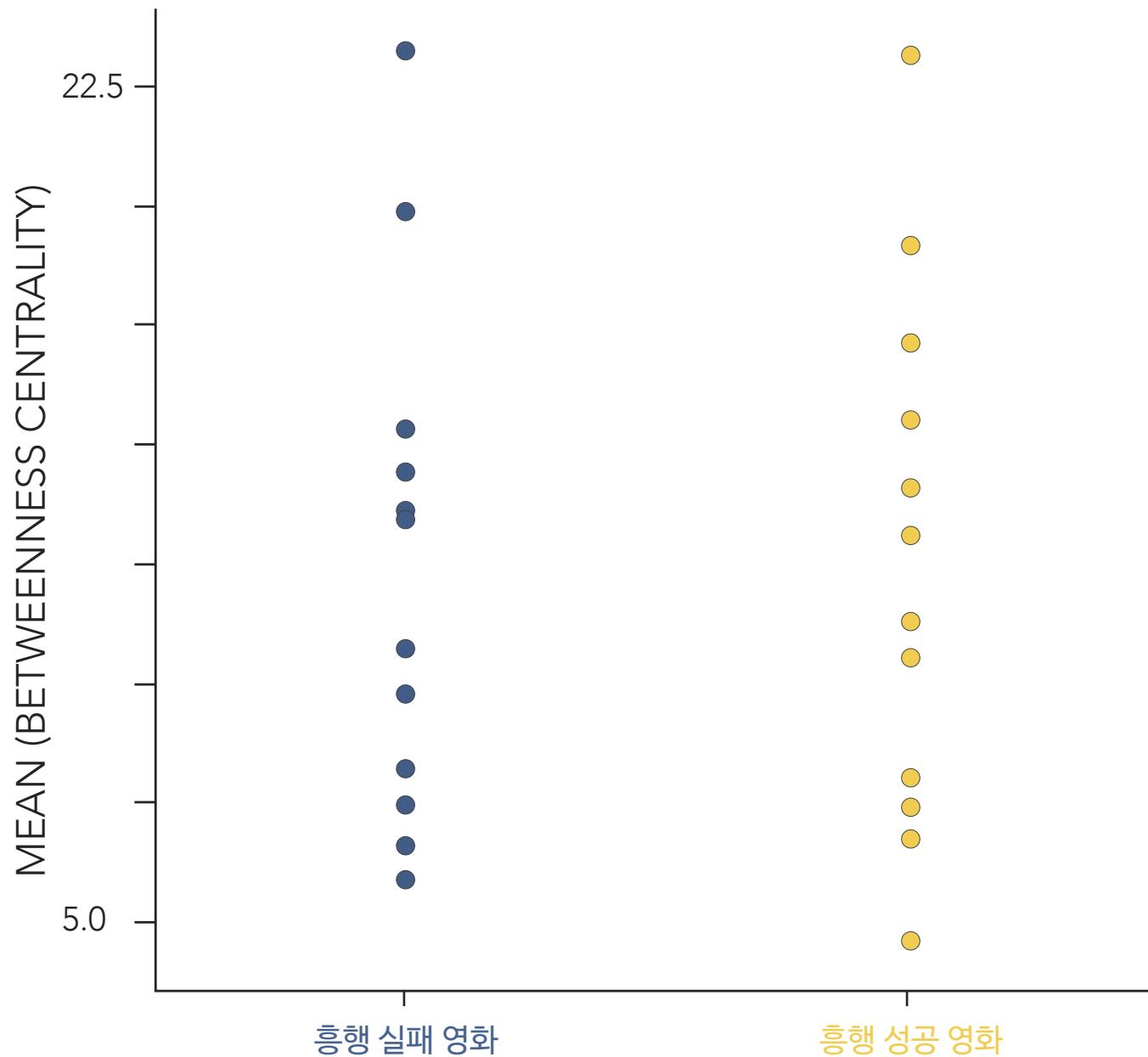
## RESULTS

통계량 시각화 결과



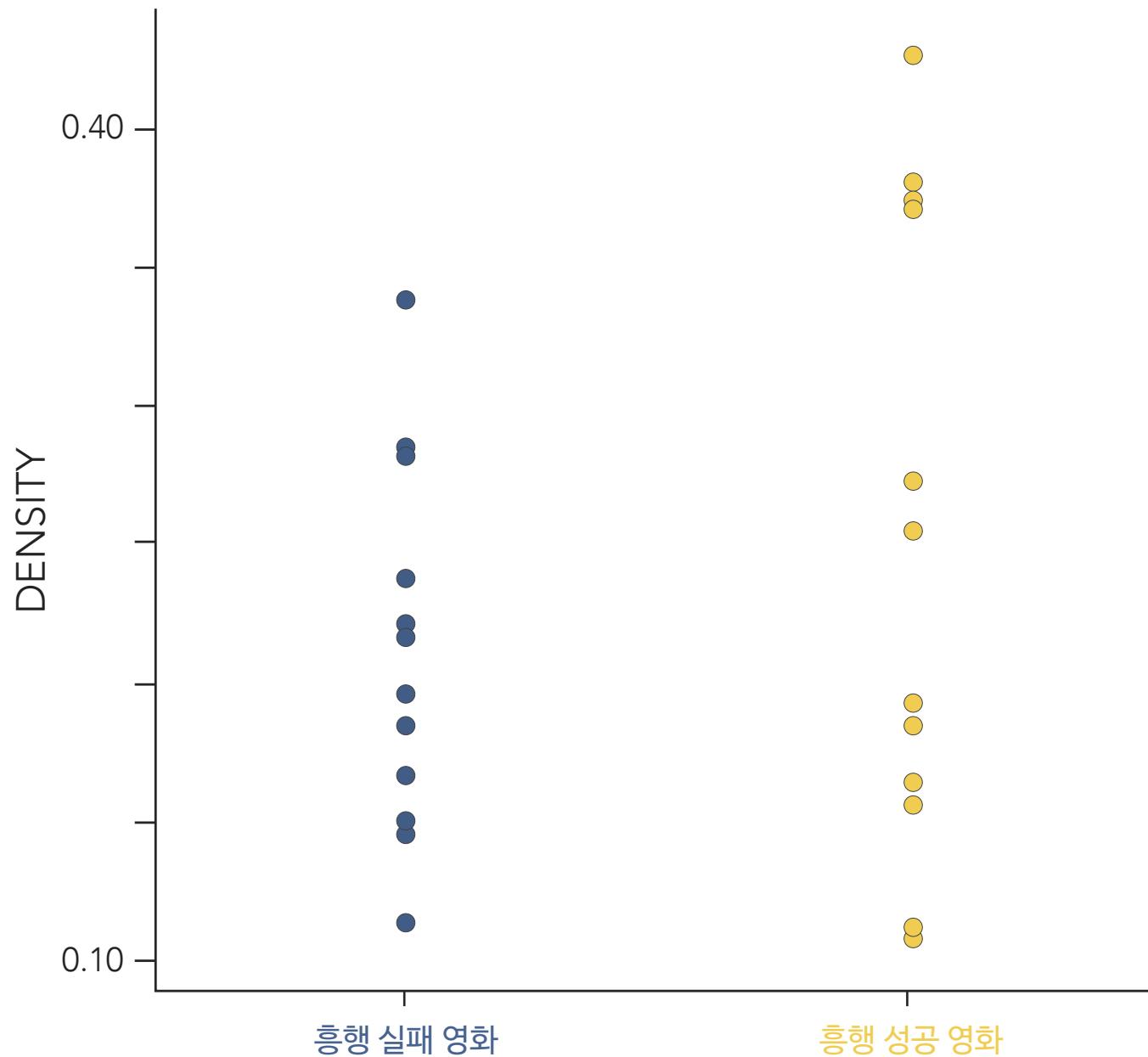
## RESULTS

통계량 시각화 결과



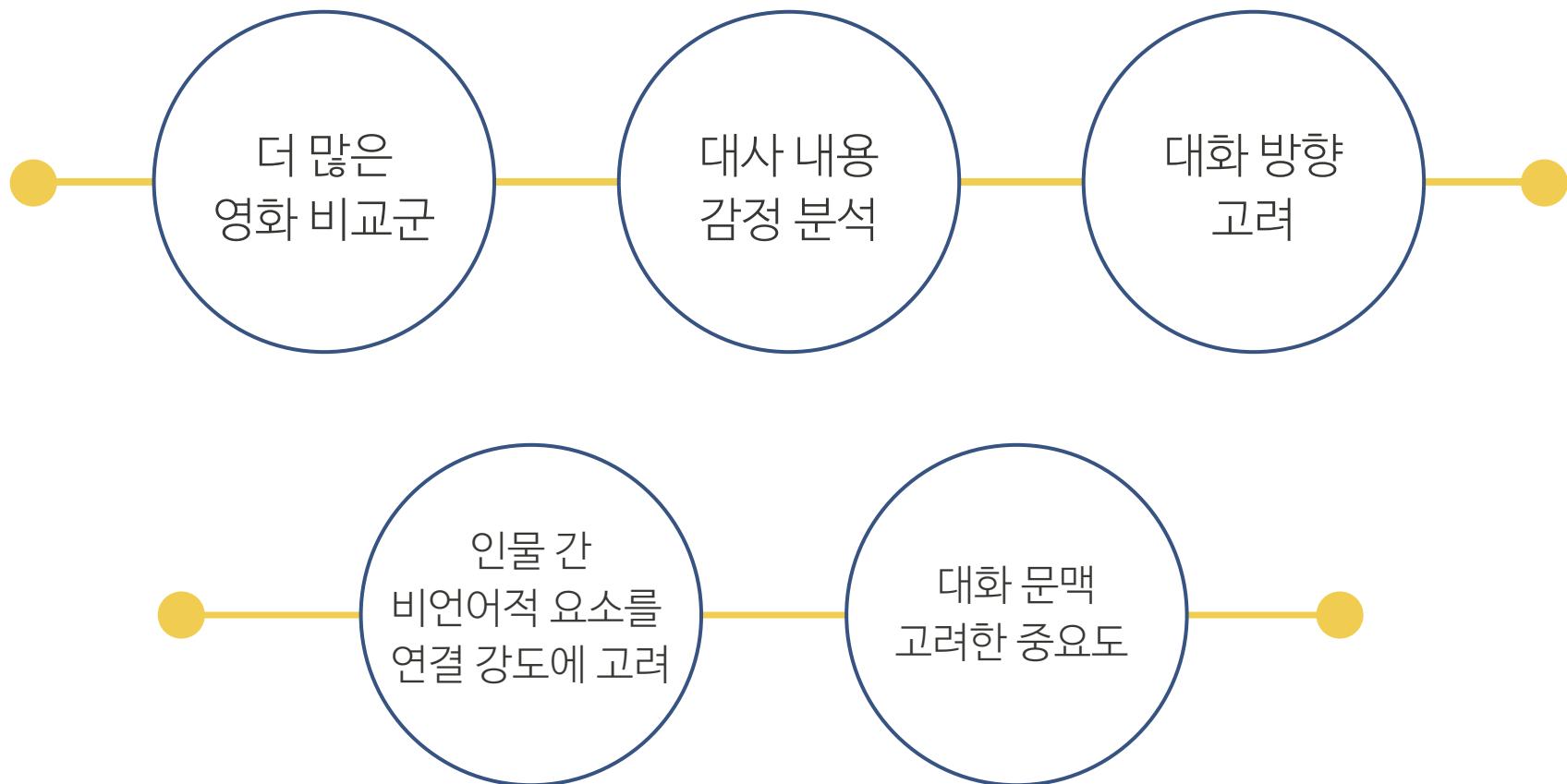
## RESULTS

통계량 시각화 결과



## CONCLUSION

다음 프로젝트 개선점



이야기의 네트워크 구조가 흥행에  
영향을 미친다고 보기

이야기의 네트워크 구조가 흥행에  
영향을 미친다고 보기 어렵다.

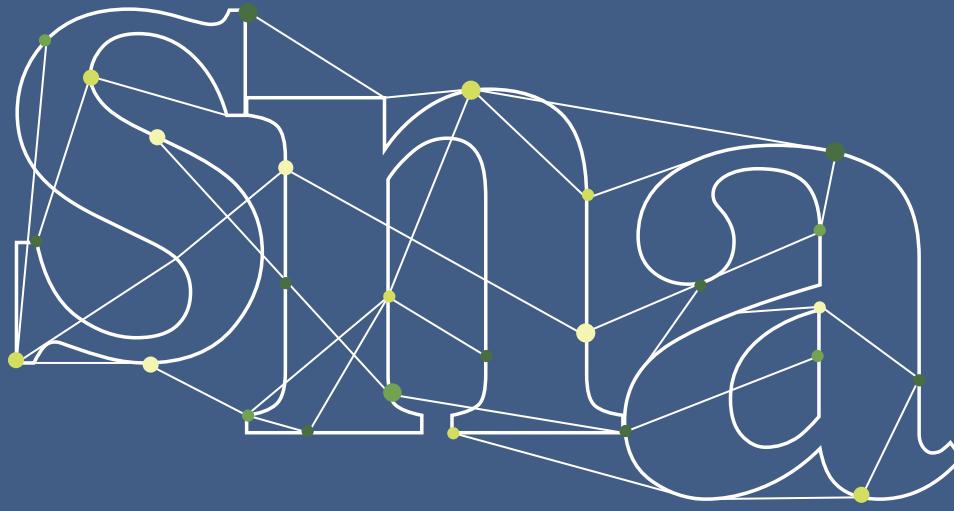
이야기의 네트워크 구조가 흥행에  
영향을 미친다고 보기 어렵다.

다만, 다소 완화된 기준 하에서  
**mean degree** 값의 차이는 의심해볼 수 있다!

이야기의 네트워크 구조가 흥행에  
영향을 미친다고 보기 어렵다.

다만, 다소 완화된 기준 하에서  
**mean degree** 값의 차이는 의심해볼 수 있다!

: 다른 인물과의 연결 횟수  
인물간 연결이 긴밀한 영화가 흥행 가능성은 높을 '수도' 있다.



감사합니다! Q&A