



# 수업계획서 텍스트마이닝 결과를 기반으로한 네트워크형 강의 검색 시스템

2017. 2. 17.

박선주	경영대학 경영학과 교수
유건식	경영대학 경영학과 석사과정
김택희	LikeLion 개발자



“... 연세대학교는 네트워크 시대의 창의력을 키우기 위해 다양한 프로그램을 개발 운영하고 있습니다. **현대의 지식은 서로 이질적인 여러 지식이 연결되고 모이면서 완전히 새로운 지식을 만들어내고 있습니다.** ... 학문의 분과영역을 뛰어넘어 타 분야 연구자와 대화할 수 있는 능력을 배양해야 할 것입니다. 이를 통해 지능(intelligence)의 향상은 물론이고 기존의 스마트한 아이디어를 연결하여 새로운 것을 만들어내는 외지능(Extelligence)을 키울 수 있을 것으로 생각합니다. 창의력이란 무에서 유를 만들어내는 능력이 아니라, 이미 있는 것들을 연결시켜 새로운 것을 창조하는 ‘생각의 네트워킹’ 능력이기 때문입니다. ...”

- 연세대학교 18대 총장 김용학

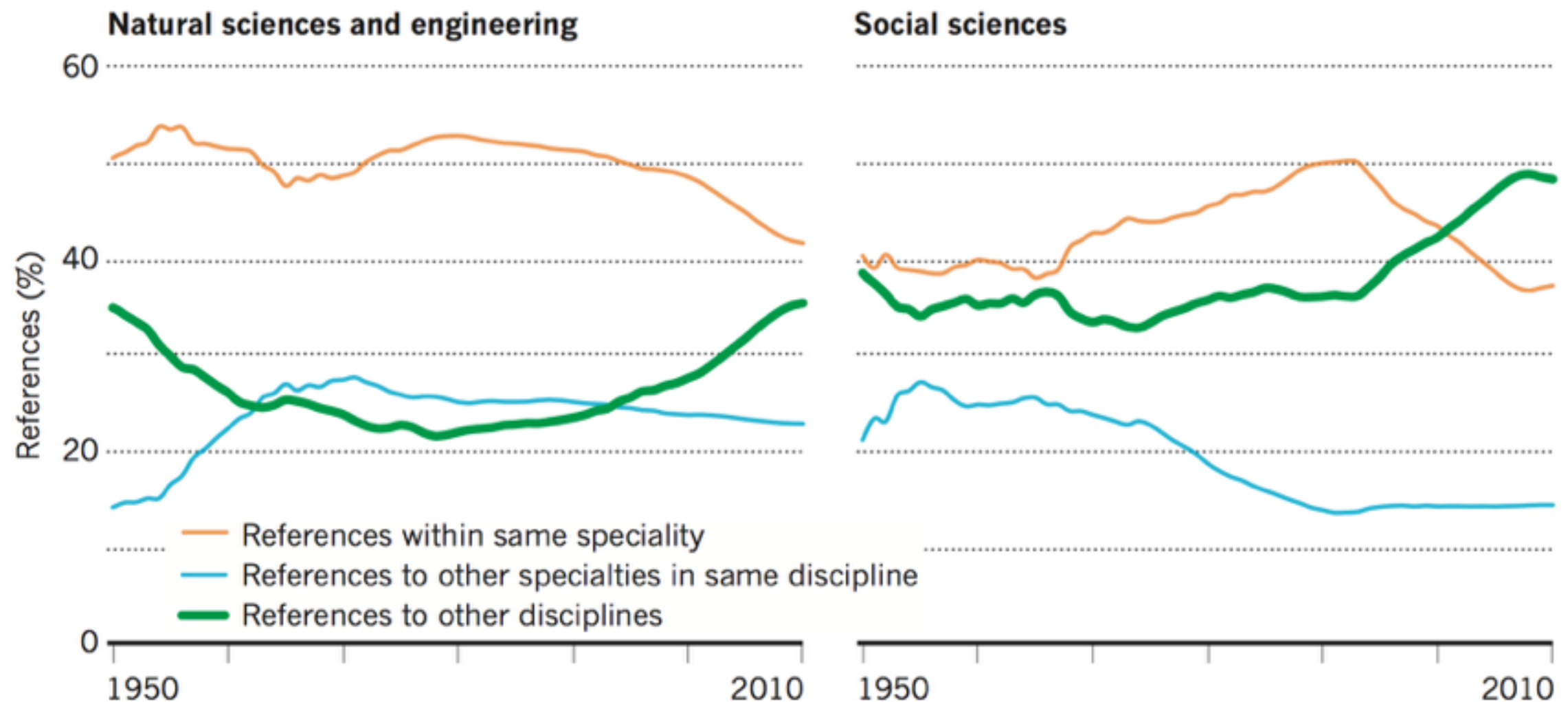
# 목 차

1. 연구 동기
2. 연구 문제
3. 연구방법론
4. 연구 결과
5. 논의점 및 향후 연구

# 1. 연구 동기

## (1) 학계의 연구 동향

- Van Noorden, R. (2015). Interdisciplinary research by the numbers. Nature.

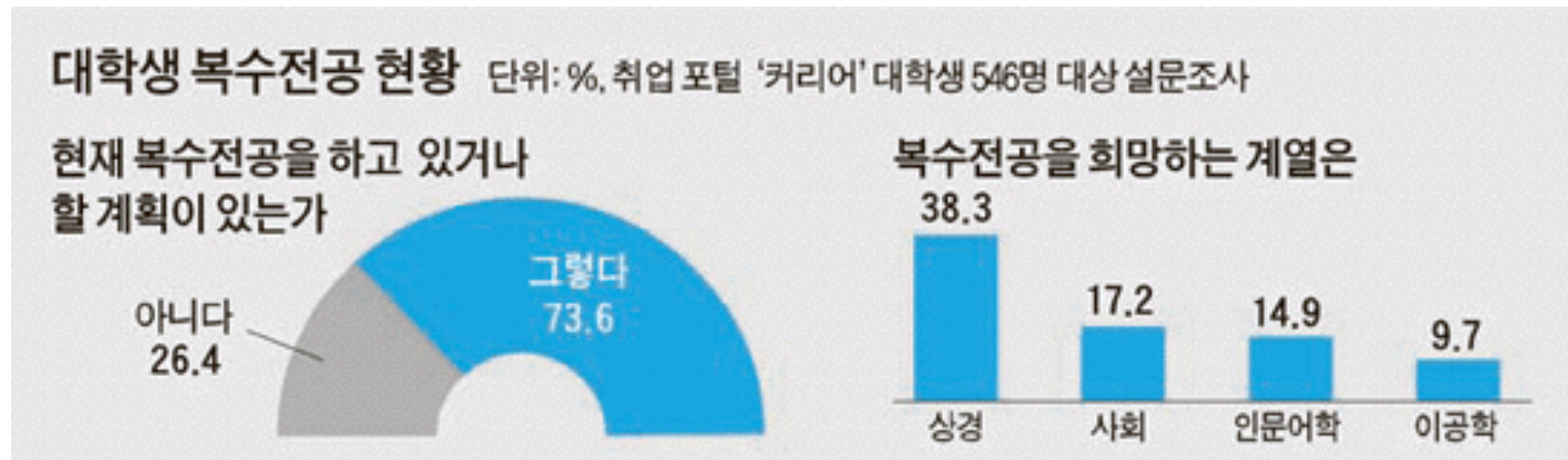


- 김용학, 윤정로, 조혜선, & 김영진. (2007). 과학기술 공동연구의 연결망 구조. 한국사회학.

# 1. 연구 동기

## (2) 학부 학생들의 동향

· 중앙일보 2011.11.8



- Russell, A. W., Dolnicar, S., & Ayoub, M. (2008). Double degrees: double the trouble or twice the return?. Higher Education.

Response	Percent
I gain a broader base of knowledge	76
It offers career/work opportunities	72
I acquire diverse skills	69
It is interesting	67
I can transfer knowledge and/or skills between disciplines	64
It is challenging	50
I have contact with diverse perspectives and ideas	47

Response	Percent
Difficulties with timetable or course structure	61
Different faculties have differing expectations	53
Long duration of the double degree	44
Heavy workload	40
Lack of administrative and educational support	39
Lack of integration or linking between your disciplines	33
Lack of recognition of your achievements as a double degree student	31
Lack of a social peer group	27
A sense of not belonging	18



## 2. 연구 문제



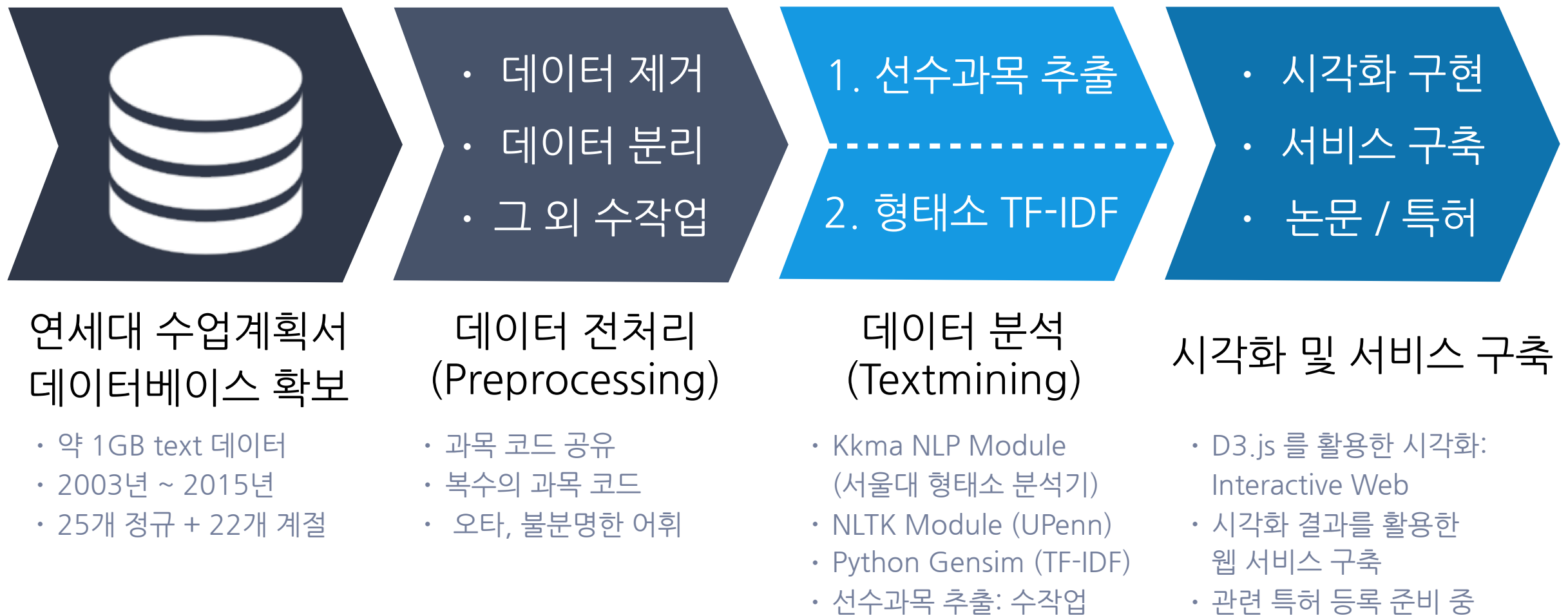
학부 수준에서 융합형 교육은 어떻게 이루어지고 있는가?

1. 전공 지식은 얼마나 확실히 분할되어 있는가?
2. 전공 간 융합 접점은 어떻게 발견할 수 있는가?
3. 현재 편람 검색 시스템보다 융합형 교육을 지원하기에 알맞는 시스템은?
4. 학생 개개인의 융합적 학업 성취는 어떻게 표현할 수 있을까?

Response	Percent
Difficulties with timetable or course structure	61
Different faculties have differing expectations	53
Long duration of the double degree	44
Heavy workload	40
Lack of administrative and educational support	39
Lack of integration or linking between your disciplines	33
Lack of recognition of your achievements as a double degree student	31
Lack of a social peer group	27
A sense of not belonging	18

### 3. 연구 방법론

#### 연구 프로세스





### 3. 연구 방법론

---

#### 확보 데이터

- A. 연세대학교 강의계획서
- B. 13년 전수조사(의대, 약대, 간호대, 음대 제외\*)  
\* 의대, 약대, 간호대는 이중전공이 불가능
- C. 2003년 ~ 2015년(25개 봄/가을 학기 + 22개 계절학기)
- D. 총 48,960 개 수업계획서 문서
- E. 총 6,618 개 수업 학정번호  
e.g. 회계원리 BIZ1101
- F. 약 13만개의 서로다른 형태소 추출
- G. 그 외 학정번호 정보 및 단과대 정보는 제공처가 없는 관계로  
개별 수집 및 수작업으로 진행
- H. 대학 차원의 데이터 협조 미비

### 3. 연구 방법론

#### 데이터 전처리

##### A. 중복된 데이터 제거 / 분리해야하는 데이터 분리 [6618 → 4898]

예시)

CSI3107	시스템프로그래밍
DS008	시스템프로그래밍

EEE2001	주니어세미나(뇌공학)
EEE2001	주니어세미나(Computational intelligence)
EEE2001	주니어세미나(의료영상의이해와MATLAB코딩)
EEE2001	주니어세미나(Matlab이용한의료영상처리입문)

처리 근거가 부족한 경우: 학정번호와 제목이 모두 다르지만, 사실상 같음

KOR3409	국어정보학	0.924529966	MSE3320	재료의구조및결함	0.914567085
KOR4406	자료전산처리방법론		MST3480	재료결함론	

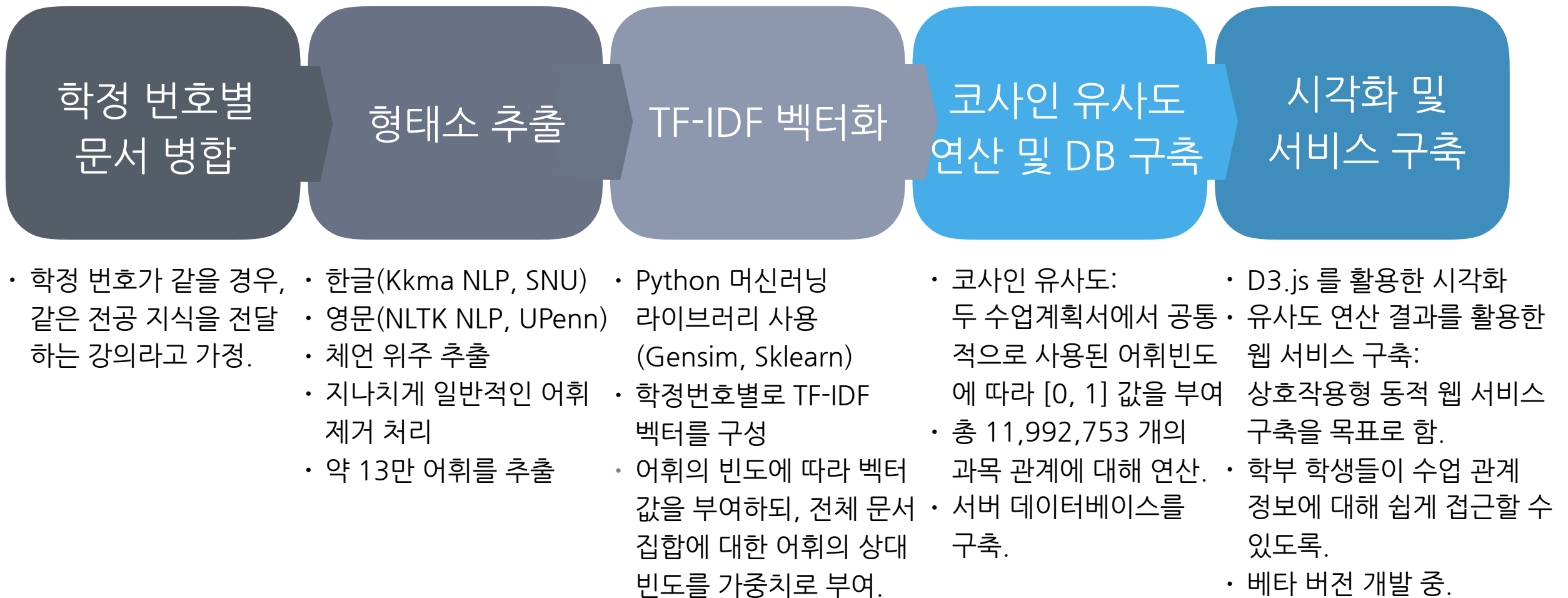
##### B. 선수과목 처리: 의미가 불분명하여, 수작업 진행

선수과목(선수학습)	STAT 3126 and STAT 3109 or equivalent courses in statistics or permission of instructor
------------	---

##### C. 그 외 정보를 왜곡하는 정보(e.g. 교수 연락처) 등을 제거(regex 응용)

### 3. 연구 방법론

#### 텍스트마이닝 / 네트워크 구축



### 3. 연구 방법론

#### 텍스트마이닝 / 네트워크 구축

유사도 연산 결과 e.g.) “회계원리(1)”

RepCode1	RepCode2	Value	
BIZ3108	BIZ1101	0.717674	회계원리(2)
BIZ3109	BIZ1101	0.680081	중급회계(1)
ENG3008	BIZ1101	0.628497	
BIZ3155	BIZ1101	0.589805	중급회계(2)
PUB4113	BIZ1101	0.564452	
BIZ4116	BIZ1101	0.517561	고급회계
BIZ4113	BIZ1101	0.490721	재무회계연습
BIZ4112	BIZ1101	0.488705	회계학연습
IEE1104	BIZ1101	0.426419	
BIZ3110	BIZ1101	0.419793	원가회계
YCJ1401	BIZ1101	0.408334	
BIZ3107	BIZ1101	0.376631	법인세법

urityTable 5

### 3. 연구 방법론

#### 텍스트마이닝 / 네트워크 구축

유사도 연산 결과 e.g.) “회계원리(1)”

RepCode1	RepCode2	Value
BIZ3108	BIZ1101	0.717674
BIZ3109	BIZ1101	0.680081
ENG3008	BIZ1101	0.628497
BIZ3155	BIZ1101	0.589805
PUB4113	BIZ1101	0.564452
BIZ4116	BIZ1101	0.517561
BIZ4113	BIZ1101	0.490721
BIZ4112	BIZ1101	0.488705
IEE1104	BIZ1101	0.426419
BIZ3110	BIZ1101	0.419793
YCJ1401	BIZ1101	0.408334
BIZ3107	BIZ1101	0.376631

ENG3008: 공학회계

PUB4113: 정부회계학

IEE1104: PRINCIPLES OF FINANCIAL ACCOUNTING

YCJ1401: 회계와 사회생활

urityTable 5

### 3. 연구 방법론

#### 텍스트마이닝 / 네트워크 구축

유사도 연산 결과 e.g.) “경영정보시스템”

RepCode1	RepCode2	Value	
UIC1753	BIZ2117	0.469561	
CTM2002	BIZ2117	0.420235	
IIE2105	BIZ2117	0.401009	
CTM2006	BIZ2117	0.355905	
BIZ3165	BIZ2117	0.355338	정보통신기술관리
BIZ3113	BIZ2117	0.350416	E-비즈니스 전략
CTM1002	BIZ2117	0.316023	
IIE3103	BIZ2117	0.315024	
ISM2108	BIZ2117	0.314287	
BIZ3112	BIZ2117	0.311517	시스템분석과 설계
CTM3001	BIZ2117	0.309007	
BIZ3177	BIZ2117	0.307426	경영 프로세스 혁신

urityTable 4

### 3. 연구 방법론

#### 텍스트마이닝 / 네트워크 구축

유사도 연산 결과 e.g.) “경영정보시스템”

RepCode1	RepCode2	Value
UIC1753	BIZ2117	0.469561
CTM2002	BIZ2117	0.420235
IIE2105	BIZ2117	0.401009
CTM2006	BIZ2117	0.355905
BIZ3165	BIZ2117	0.355338
BIZ3113	BIZ2117	0.350416
CTM1002	BIZ2117	0.316023
IIE3103	BIZ2117	0.315024
ISM2108	BIZ2117	0.314287
BIZ3112	BIZ2117	0.311517
CTM3001	BIZ2117	0.309007
BIZ3177	BIZ2117	0.307426

rityTable 4

UIC1753: INFORMATION TECHNOLOGY AND NETWORKED ORGANIZATIONS

CTM2002: INTRODUCTION TO INFORMATION SYSTEMS MANAGEMENT

IIE2105: 경영정보시스템

CTM2006: INFORMATION TECHNOLOGY AND NETWORKED ORGANIZATIONS

CTM1002: IS PROJECT MANAGEMENT

IIE3103: 정보전략계획 및 설계

ISM2108: INFORMATION SYSTEMS AND INFORMATION SOCIETY

CTM3001: BUSINESS MODELS IN CREATIVE TECHNOLOGY INDUSTRY

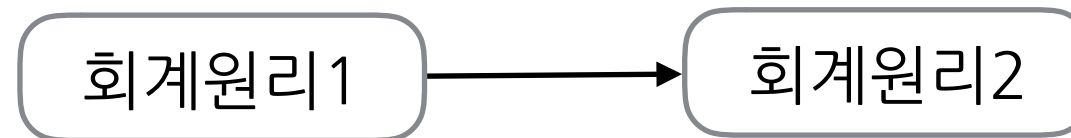


### 3. 연구 방법론

#### 텍스트마이닝 / 네트워크 구축

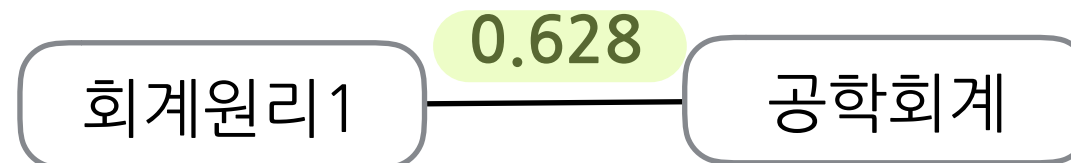
- 선수과목 관계

예시)



- 강의 간 유사성(코사인 유사도)

예시)



RepCode1	RepCode2	Value
BIZ3108	BIZ1101	0.717674
BIZ3109	BIZ1101	0.680081
ENG3008	BIZ1101	0.628497
BIZ3155	BIZ1101	0.589805

ENG3008: 공학회계

## 4. 연구 결과

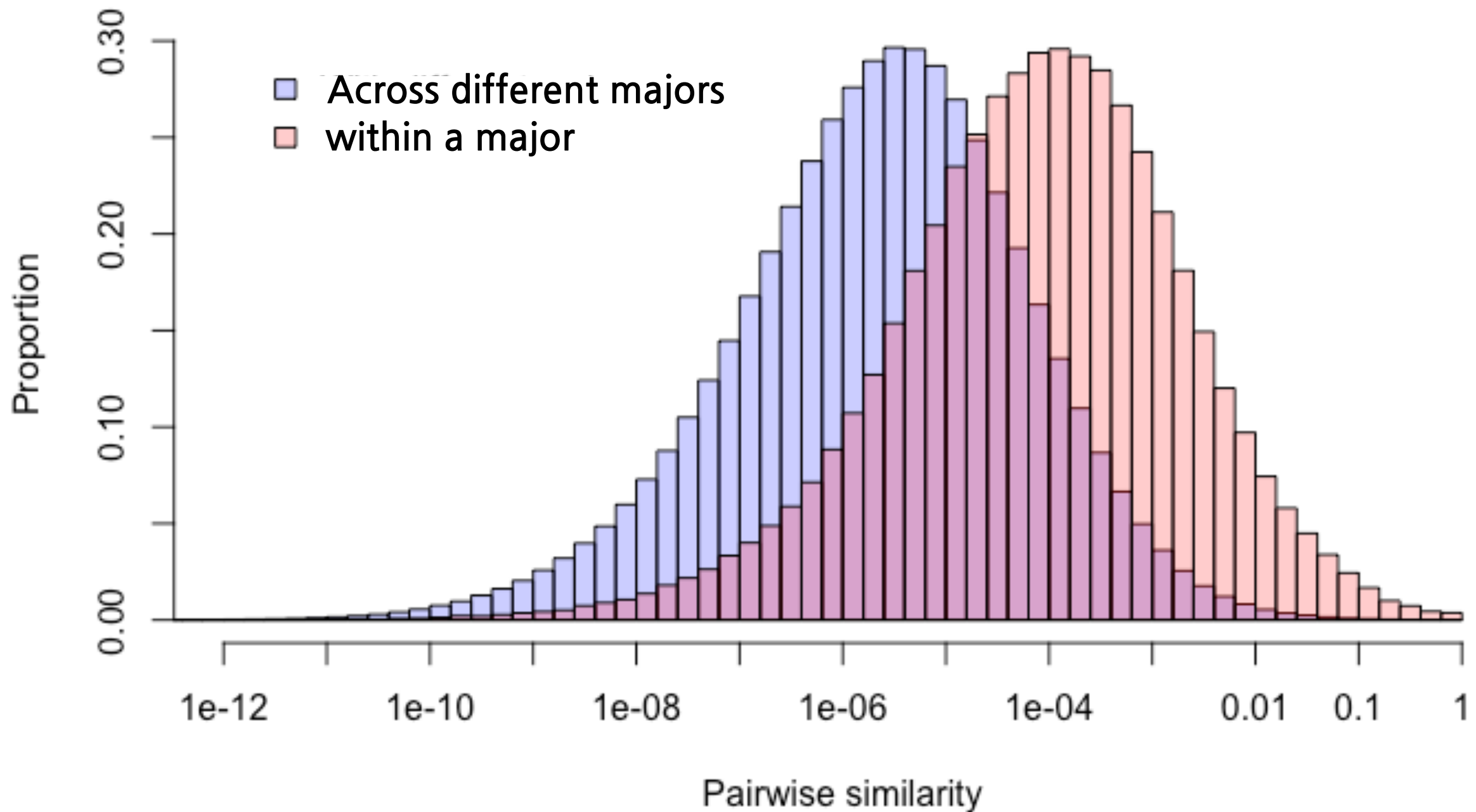
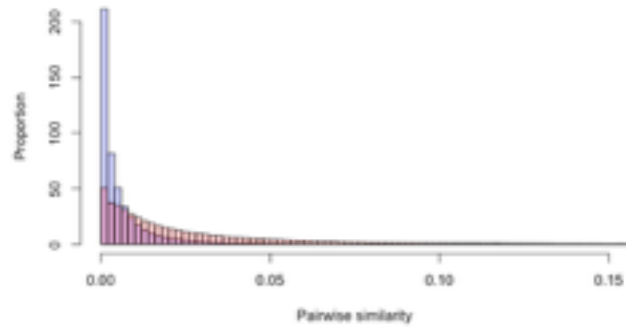
---



학부 수준에서 융합형 교육은 어떻게 이루어지고 있는가?

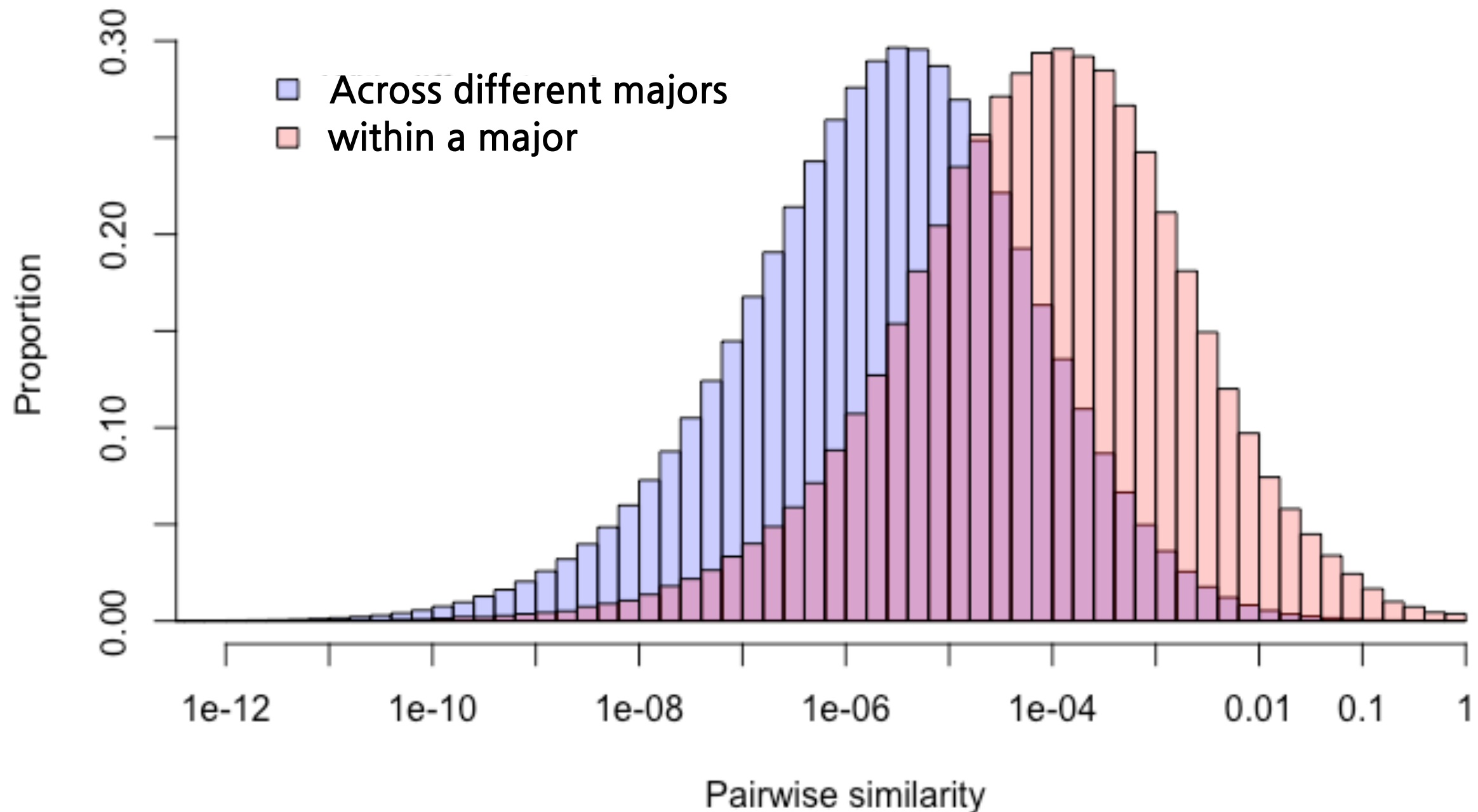
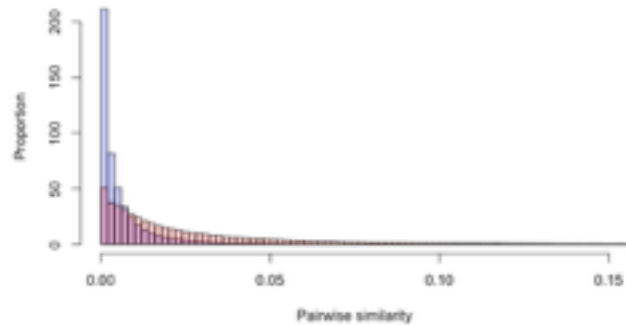
1. 전공 지식은 얼마나 확실히 분할되어 있는가?
2. 전공 간 융합 접점은 어떻게 발견할 수 있는가?
3. 현재 편람 검색 시스템보다 융합형 교육을 지원하기에 알맞는 시스템은?
4. 학생 개개인의 융합적 학업 성취는 어떻게 표현할 수 있을까?

## 4-1. 전공 지식은 얼마나 분할되어 있는가?



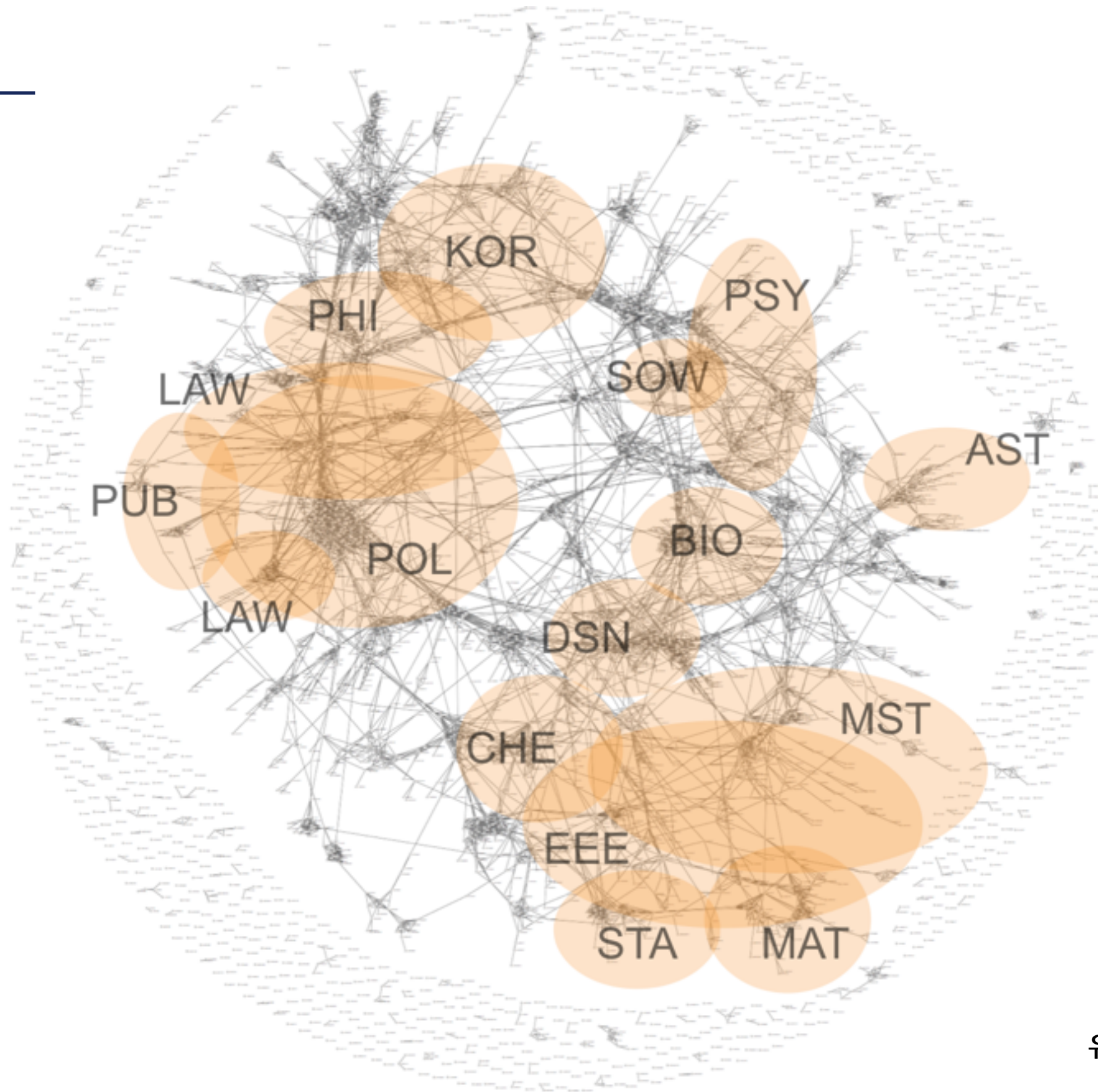
## 4-1. 전공 지식은 얼마나 분할되어 있는가?

“전공 내부의 연결이 전공 간 연결보다 강하다.  
즉, 전공 분류에 따른 커리큘럼 구축은 유의미하다.”



## 4-1. 전공 지식은 얼마나 분할되어 있는가?

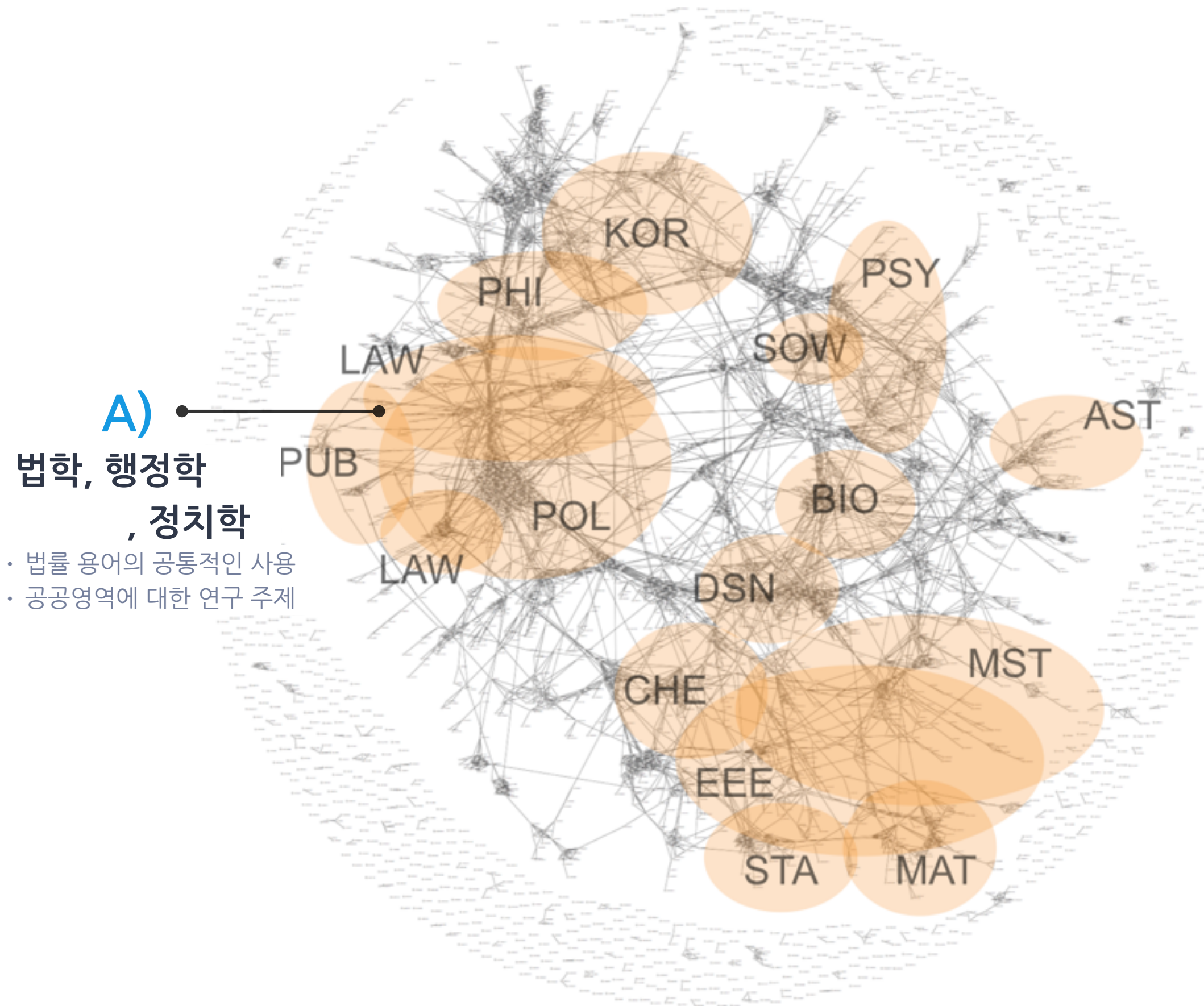
모든 전공



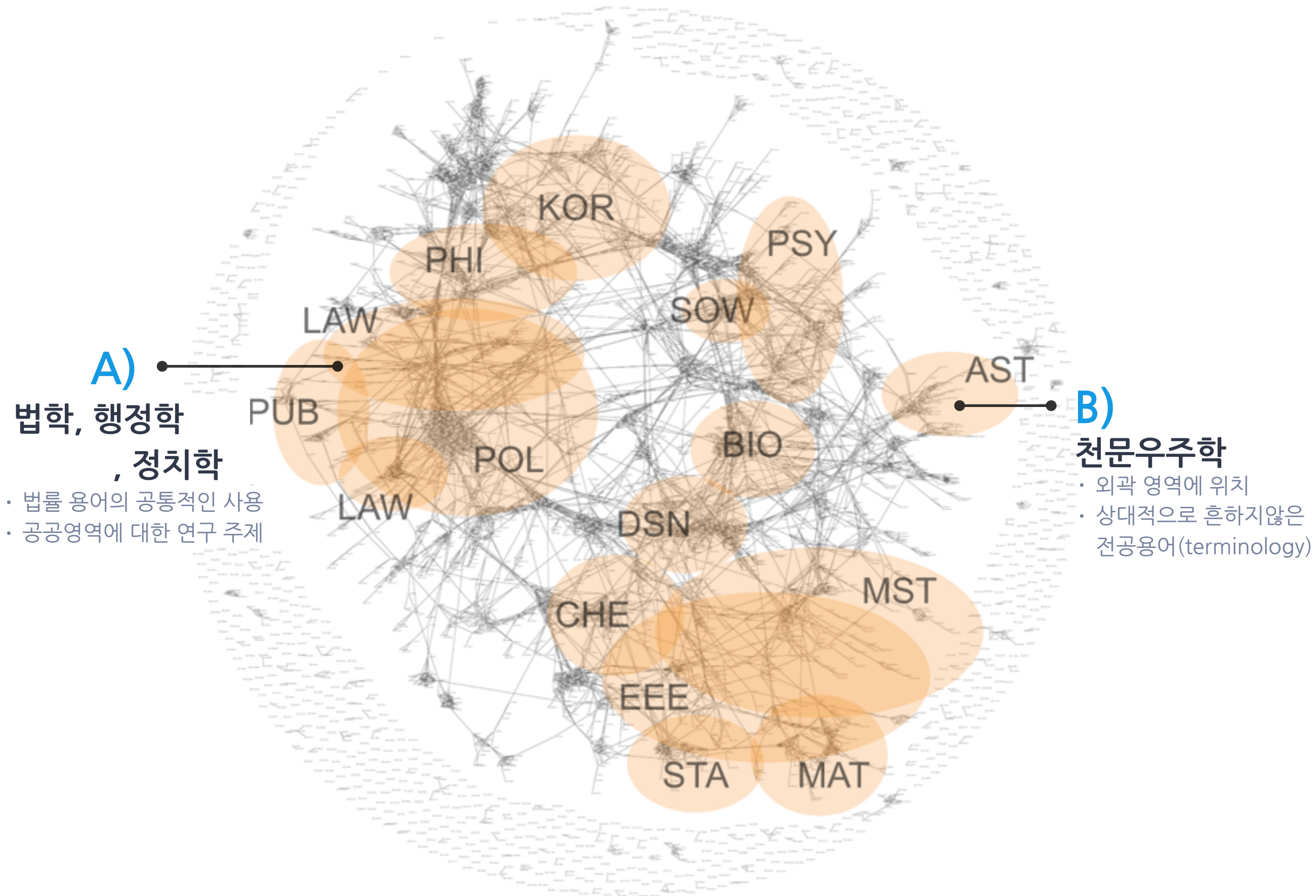
유사도 임계값  $\lambda = 0.25$



## 4-1. 전공 지식은 얼마나 분할되어 있는가?

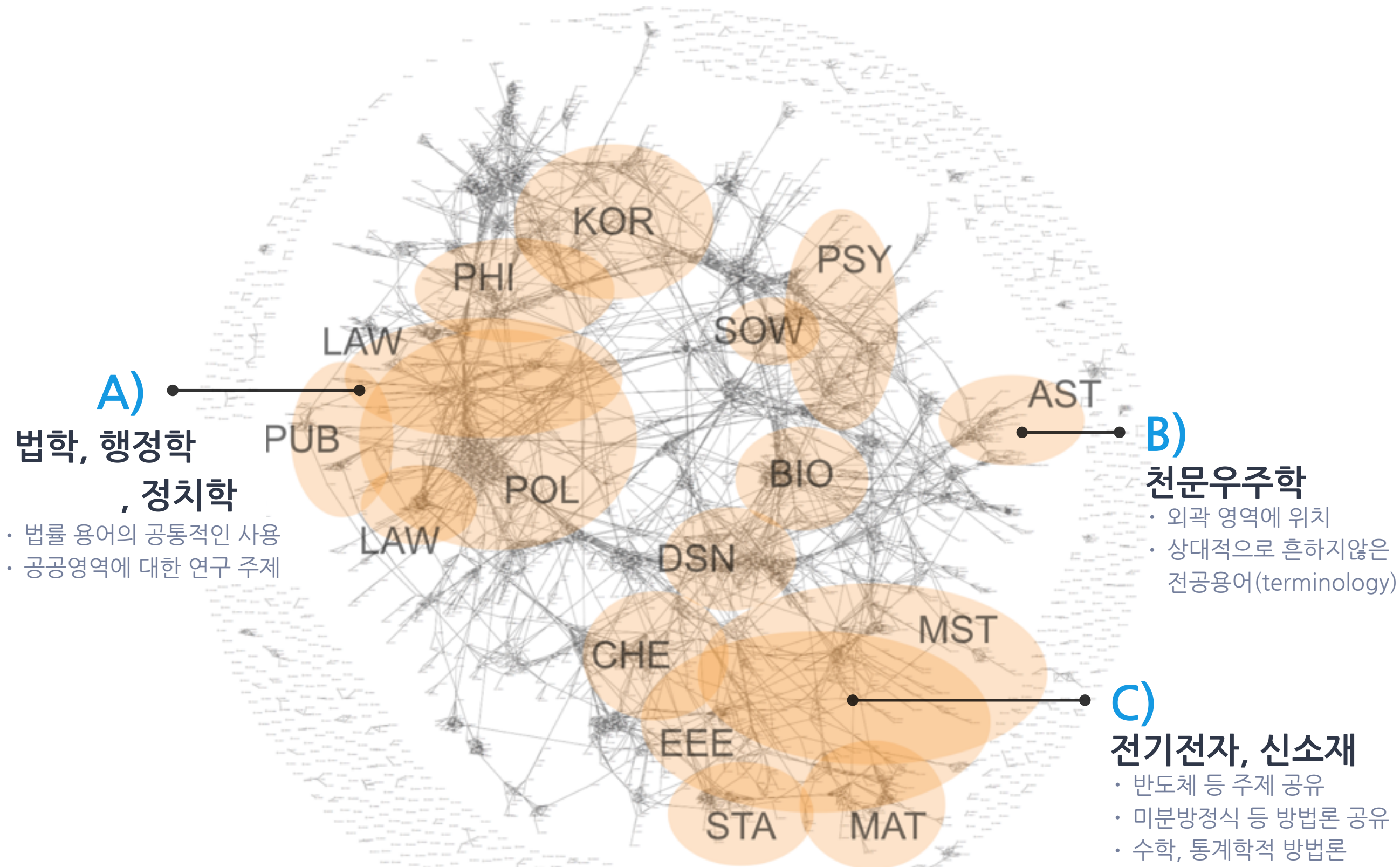


## 4-1. 전공 지식은 얼마나 분할되어 있는가?

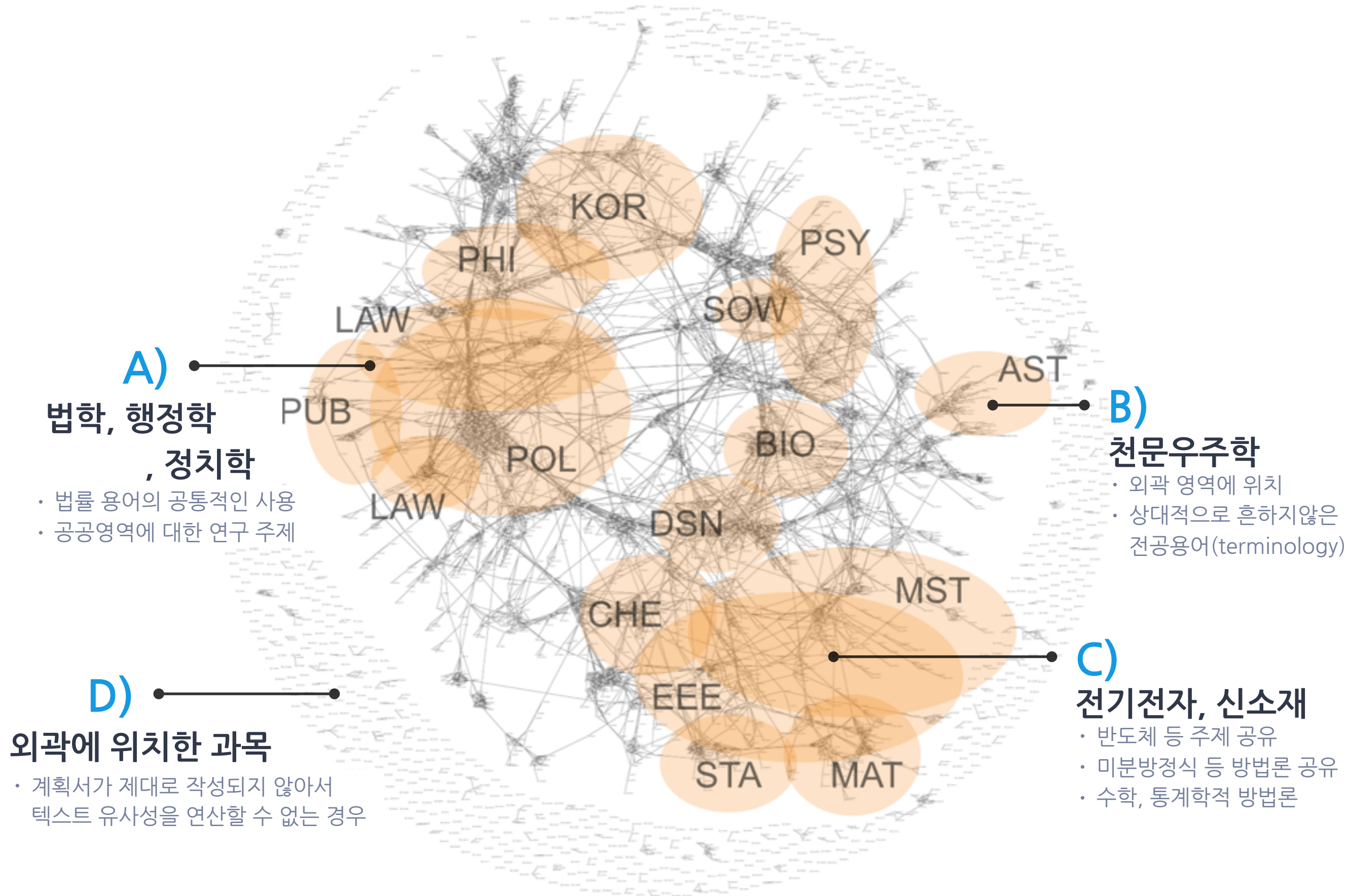




## 4-1. 전공 지식은 얼마나 분할되어 있는가?



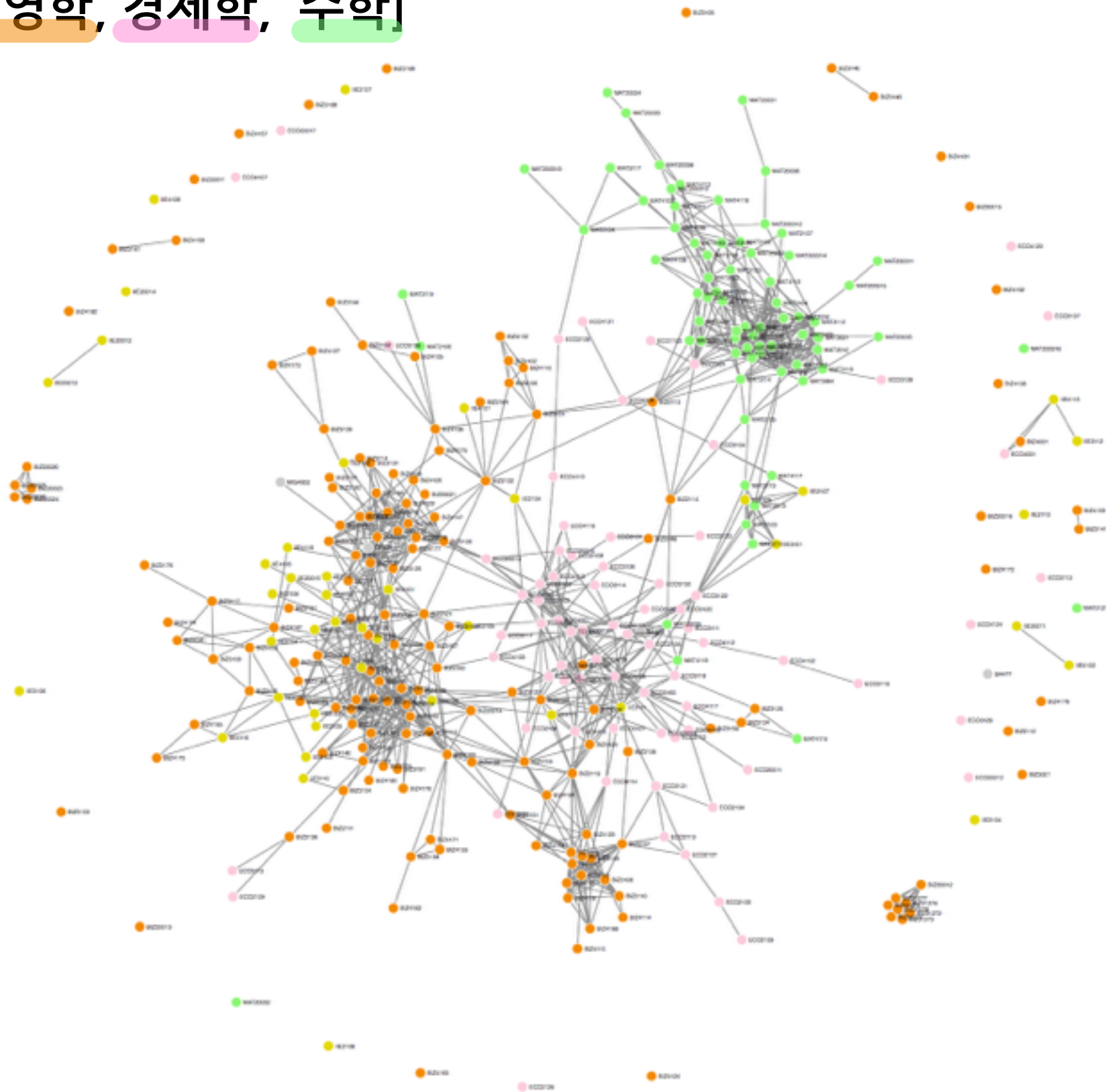
## 4-1. 전공 지식은 얼마나 분할되어 있는가?





## 4-2. 전공 간 융합 접점은 어떻게 발견할 수 있는가?

[산업공학, 경영학, 경제학, 수학]



유사도 임계값  $\lambda = 0.15$

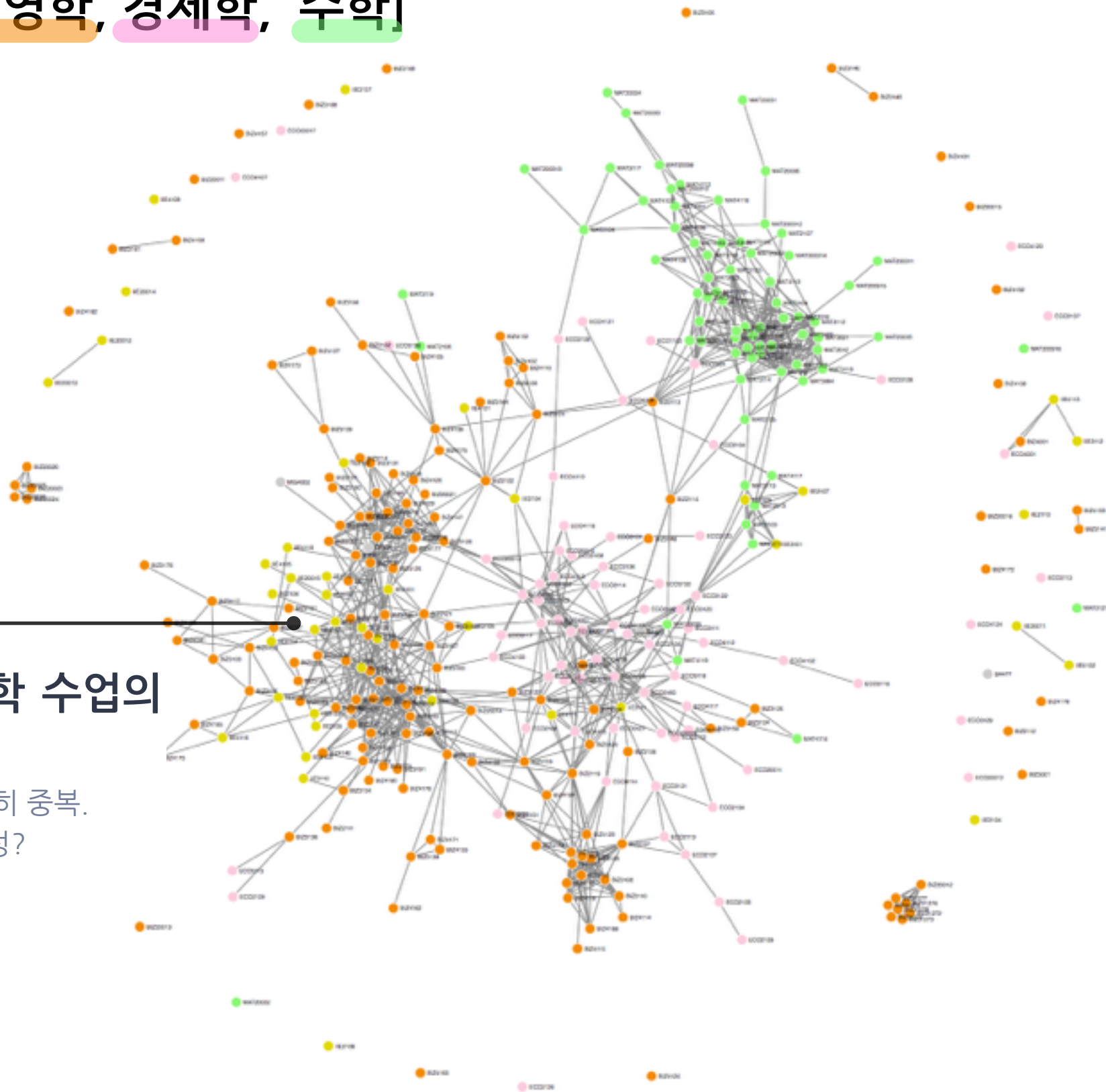
## 4-2. 전공 간 융합 접점은 어떻게 발견할 수 있는가?

[산업공학, 경영학, 경제학, 수학]

A)

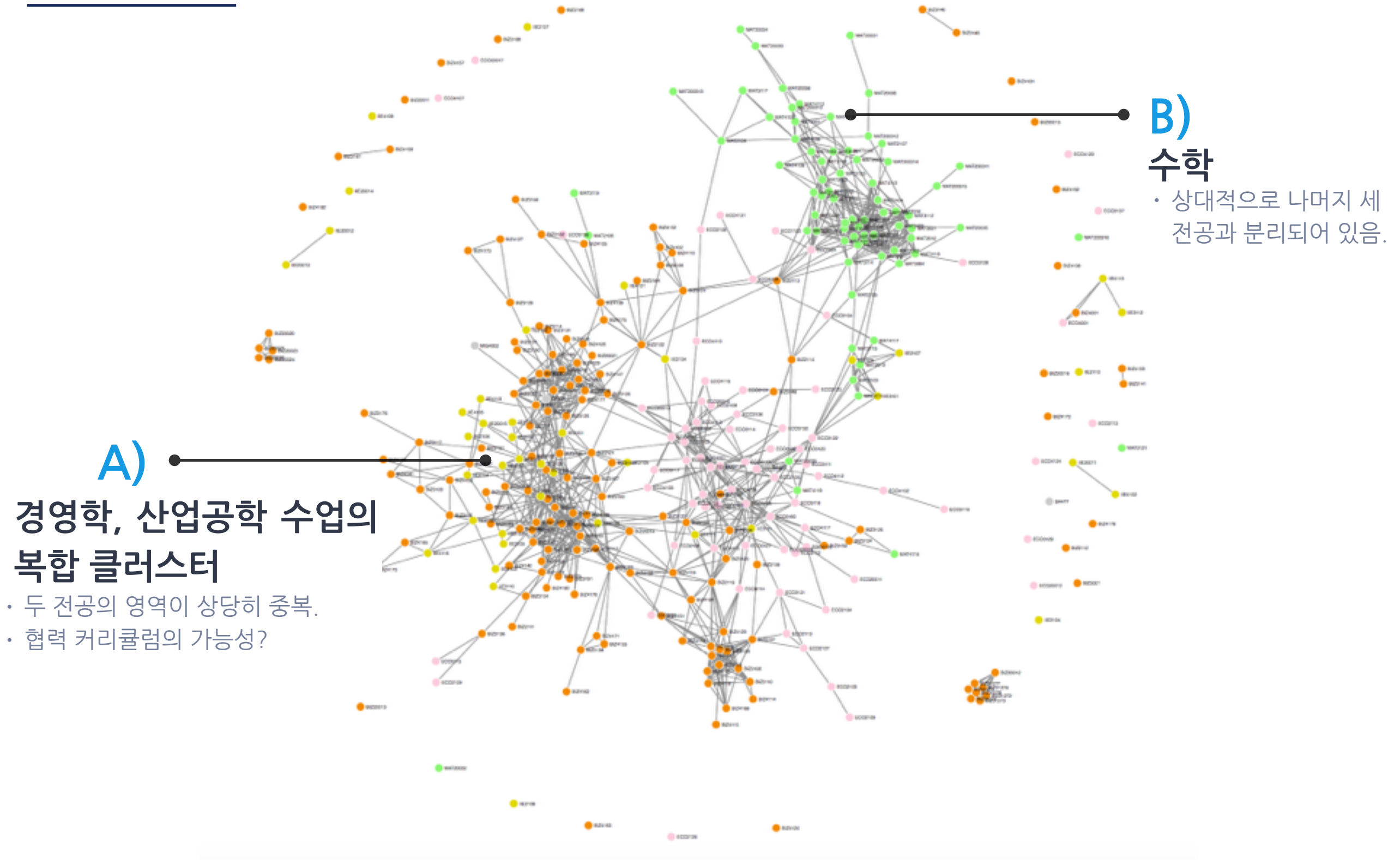
경영학, 산업공학 수업의  
복합 클러스터

- 두 전공의 영역이 상당히 중복.
- 협력 커리큘럼의 가능성?



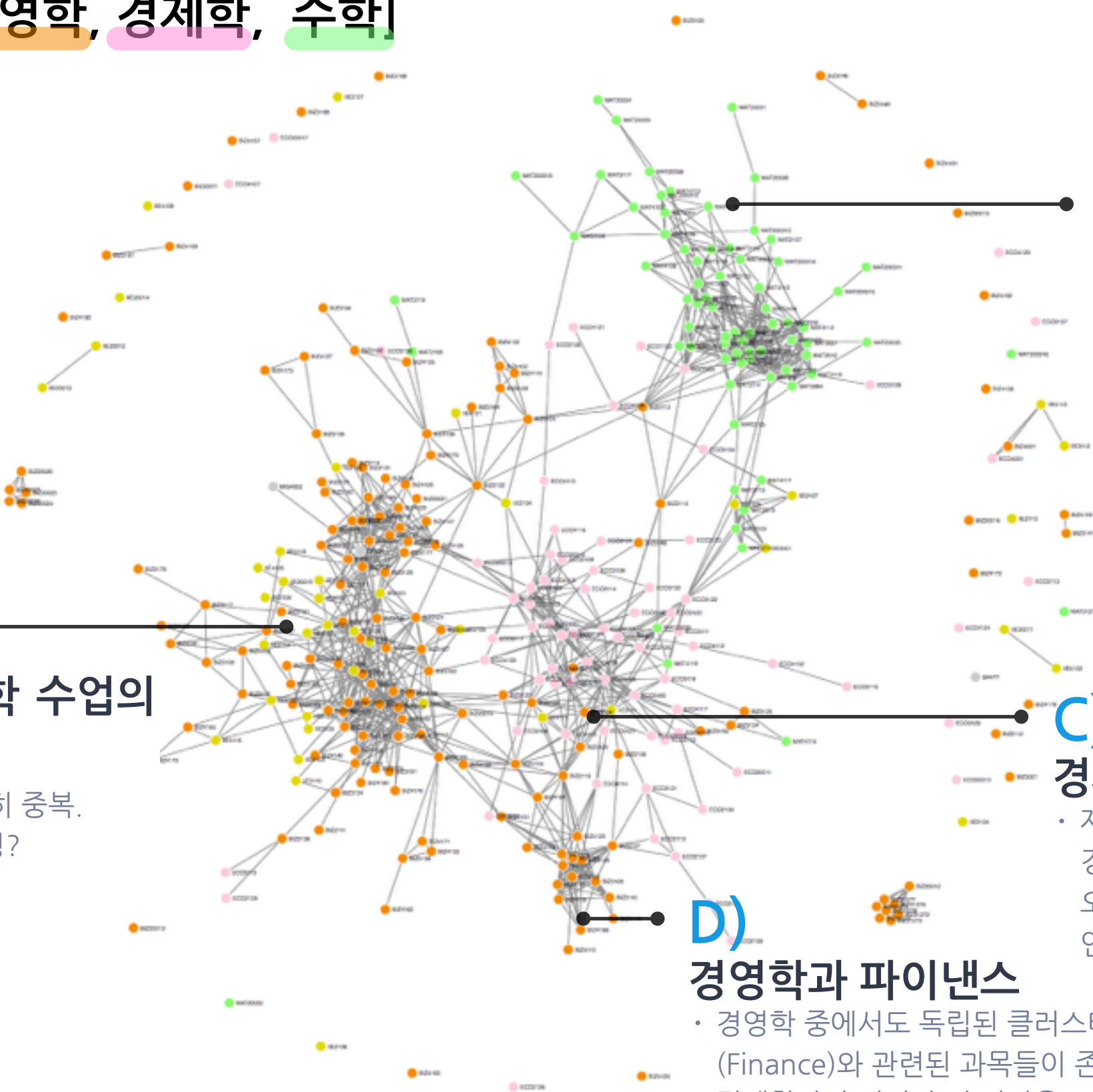
## 4-2. 전공 간 융합 접점은 어떻게 발견할 수 있는가?

[산업공학, 경영학, 경제학, 수학]



## 4-2. 전공 간 융합 접점은 어떻게 발견할 수 있는가?

[산업공학, 경영학, 경제학, 수학]



A)

경영학, 산업공학 수업의  
복합 클러스터

- 두 전공의 영역이 상당히 중복.
- 협력 커리큘럼의 가능성?

B)

수학

- 상대적으로 나머지 세 전공과 분리되어 있음.

C)

경제학과 파이낸스

- 재무(Finance)와 관련된 경영학, 산업공학 관련 과목은 오히려 경제학 과목들과 연결이 더 강함.

D)

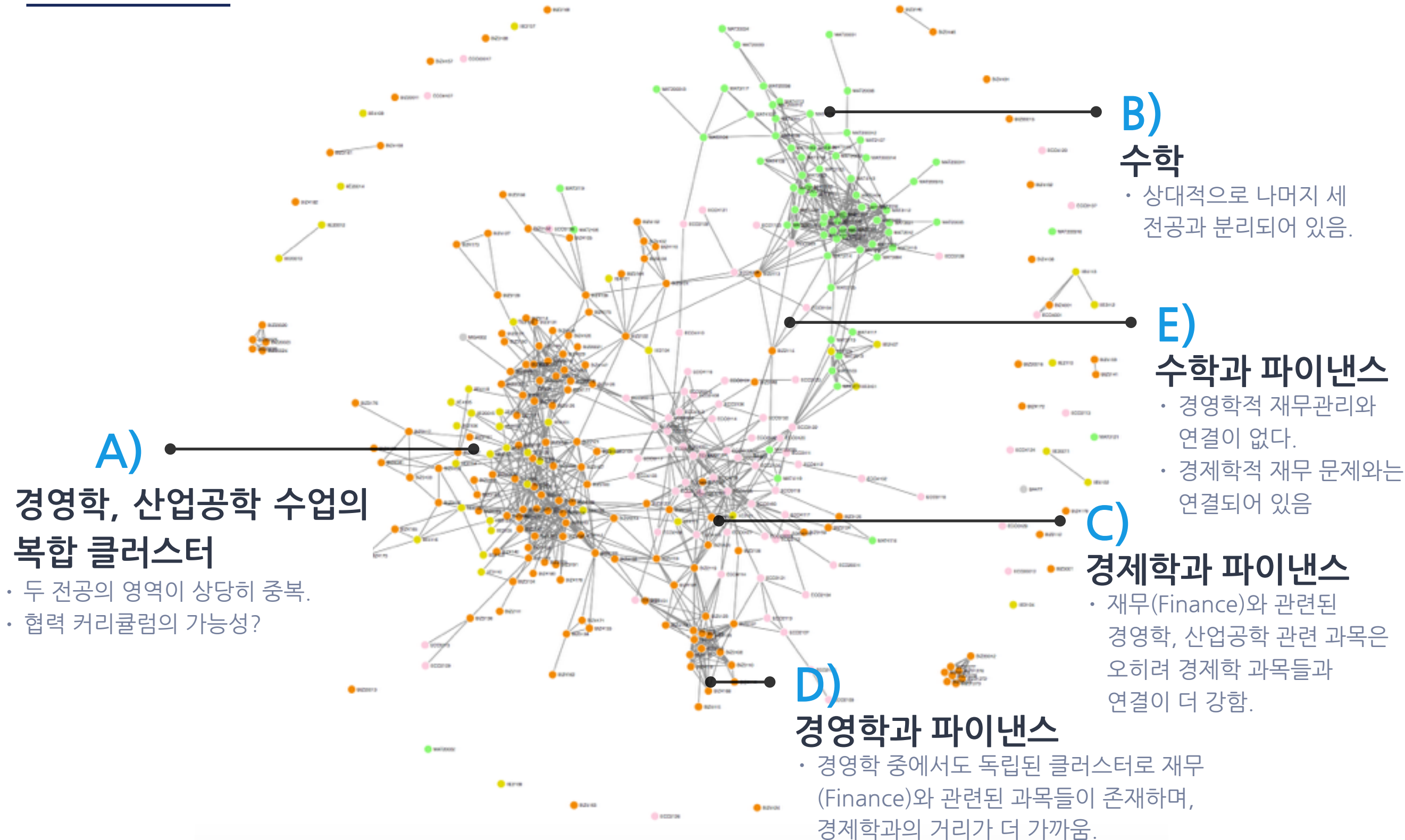
경영학과 파이낸스

- 경영학 중에서도 독립된 클러스터로 재무(Finance)와 관련된 과목들이 존재하며, 경제학과의 거리가 더 가까움.



## 4-2. 전공 간 융합 접점은 어떻게 발견할 수 있는가?

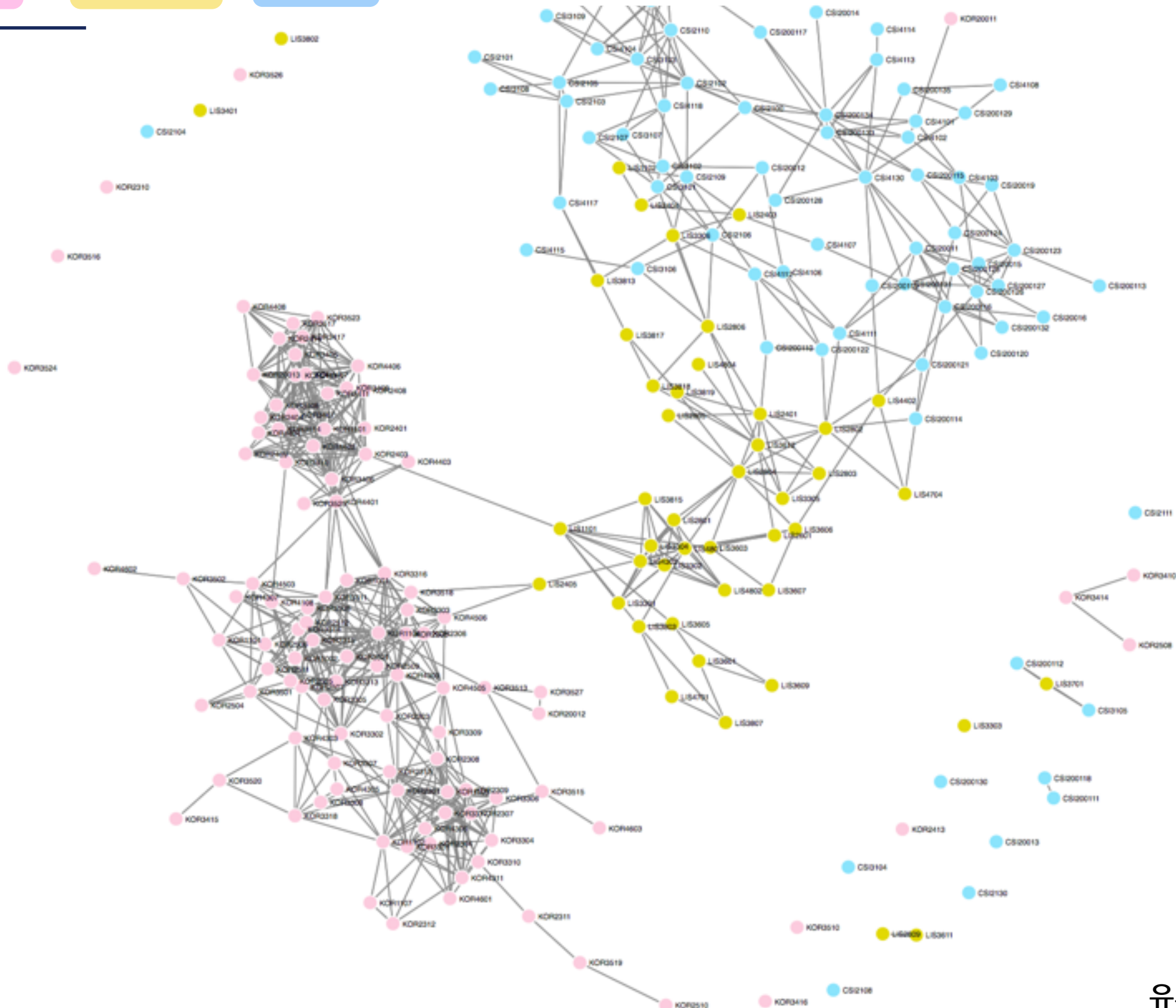
[산업공학, 경영학, 경제학, 수학]





## 4-2. 전공 간 융합 접점은 어떻게 발견할 수 있는가?

[국어국문, 문헌정보, 컴퓨터]



유사도 임계값  $\lambda = 0.15$

# 4-3. 학생들은 수강 정보를 어떻게 얻는가?

## 현행 검색 시스템

개설전공

학부

공과대학

컴퓨터과학전공

해당학기 영어강의 전체검색 :

결과내 영어강의만 보기 : ☐

전체검색

학정번호

학년도학기

2016

1학기

동일교과목 조회

교과목 개요 조회

수업계획서 조회

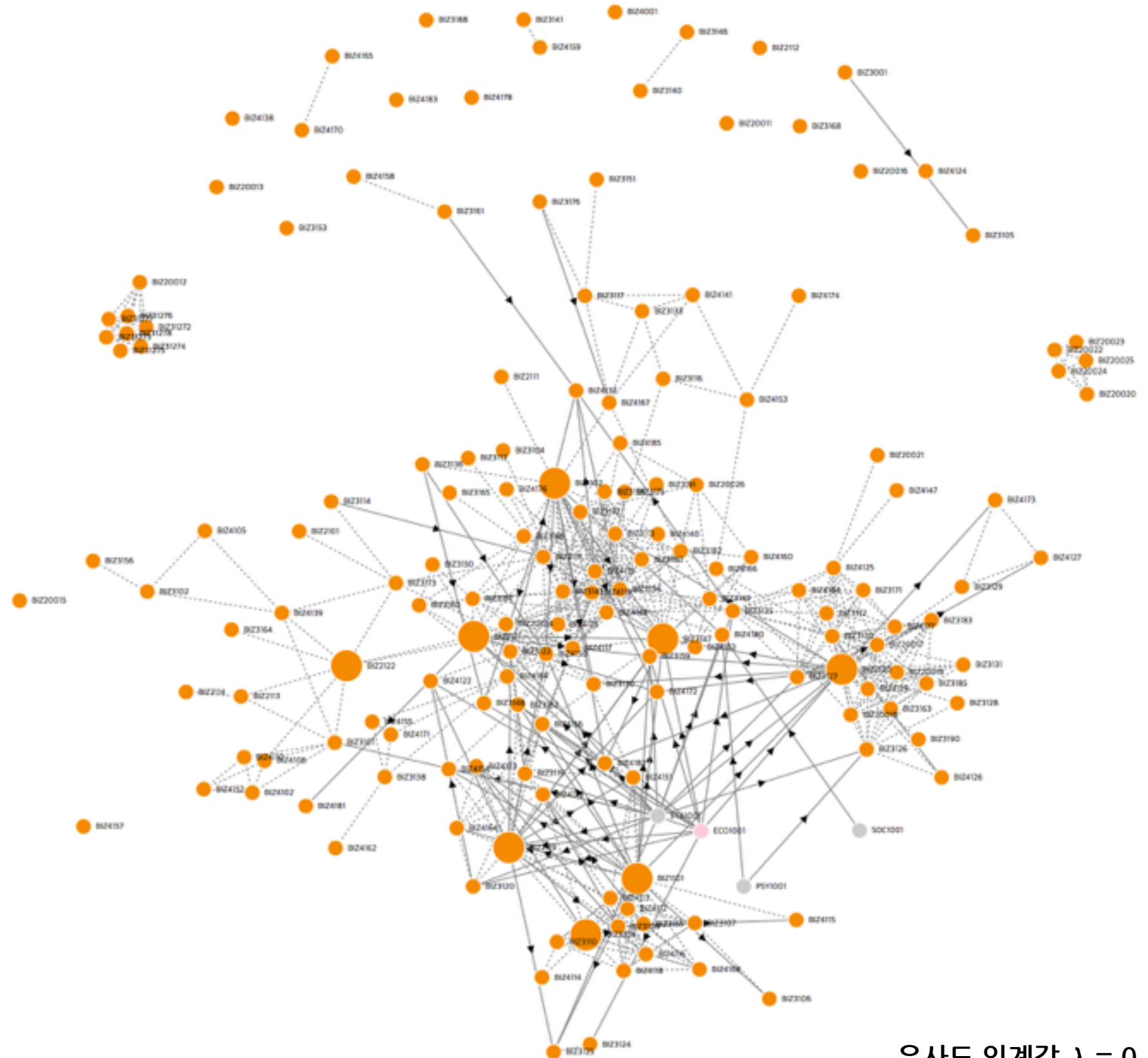
영문교과목명 조회

마일리지 수강신청 결과

학년	종별	단위	학정번호-분반(-실습)	학점	교과목명	담당교수	강의시간	강의실	유의사항
1	전필	2000	CSI2102-01-00	3	객체지향프로그래밍	김재경	월3,4,수4(수4)	공A658(공B017)	⑦
1	전필	2000	CSI2102-02-00	3	객체지향프로그래밍	김재경	월8,금5,6(월8)	공A658(공B017)	⑦
1	전선	2000	CSI2090-01-00	3	컴퓨팅적사고와프로그래밍		화5,6(화9,10(화9,10))	I자B202(I자A302(I자A301))	폐강
1	전선	2000	CSI2090-02-00	3	컴퓨팅적사고와프로그래밍				폐강
1	전선	2000	CSI2100-01-00	3	컴퓨터프로그래밍	백스텔러번트	화5,6(화9,10(화9,10(화9,10))	I자B202(I자A302(I자A301(I자B202))	③②
1	전선	2000	CSI2100-02-00	3	컴퓨터프로그래밍	김선주	목5,6(화7,8(화7,8))	공A663(공A213(공B017))	②⑦
2	전기	2000	CSI2101-01-00	3	이산구조	이수경	화5,6,목7	공A528	
2	전필	2000	CSI2103-01-00	3	자료구조	양성봉	월7/수5,6	공A542/공A577	②
2	전필	2000	CSI2103-02-00	3	자료구조	안형찬	화9,목8,9	공A663	②
2	전선	2000	CSI2001-01-00	1	주니어세미나				폐강
2	전선	2000	CSI2001-02-00	1	주니어세미나				폐강
2	전선	2000	CSI2106-01-00	3	컴퓨터과학입문	최윤철	화2,3,4	공A528	
2	전선	2000	CSI2106-02-00	3	컴퓨터과학입문	박상현	화4/목3,4	공A653/공A528	
2	전선	2000	CSI2107-01-00	3	컴퓨터시스템	김재경	수7,금2,3	공B040	
2	전선	2000	CSI2109-01-00	3	인터넷프로그래밍	이경호	화7,8,목2	공A663	

1-15 of 33

### 4-3. 융합형 교육을 지원하기에 알맞는 검색 시스템은?



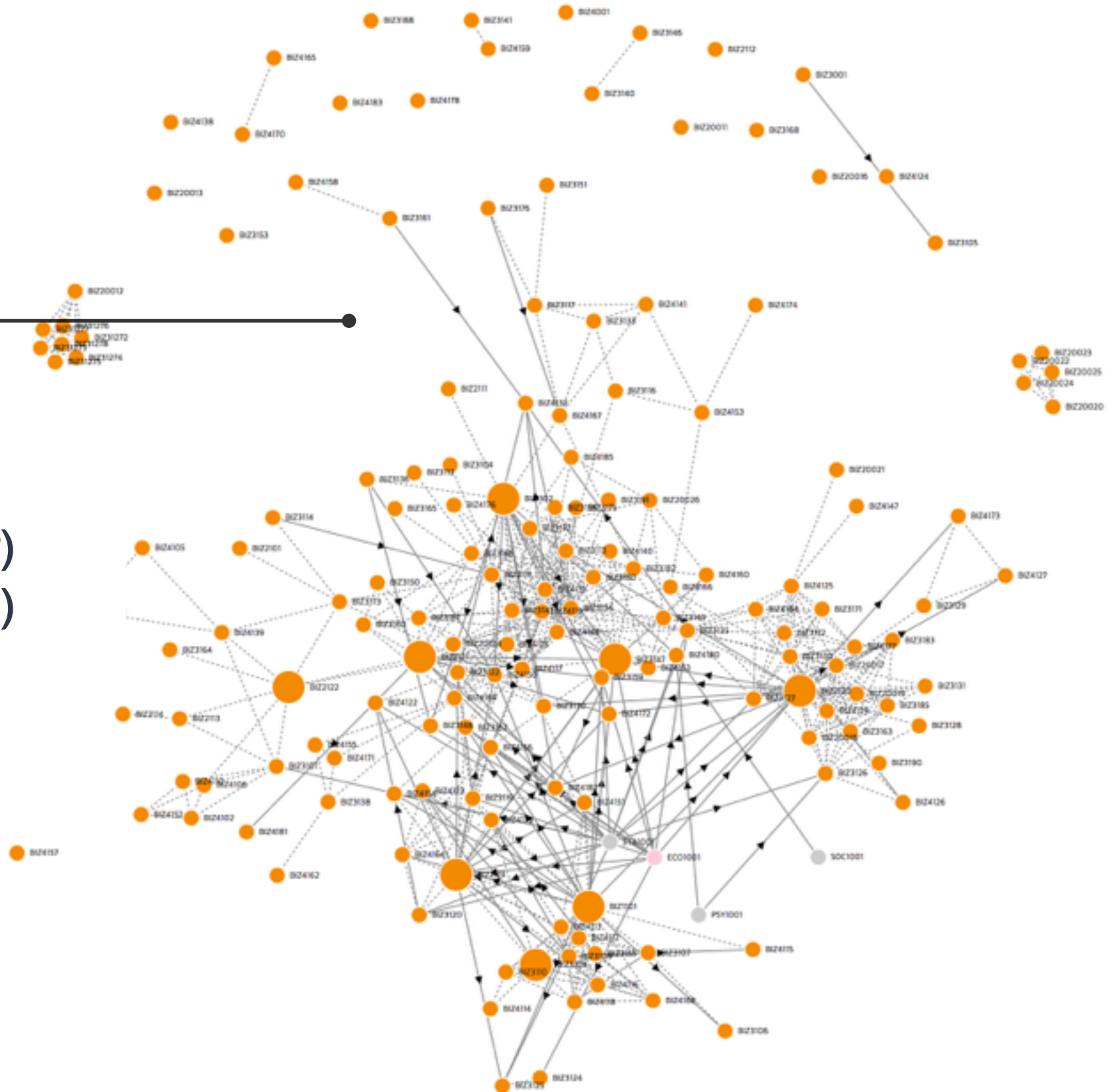
## 4-3. 융합형 교육을 지원하기에 알맞는 검색 시스템은?

[경영학]

시각화 된  
학사정보

- 필수과목(큰 원) 시각화
- 선수과목 관계 시각화(화살)
- 연관과목 관계 시각화(점선)

자연스럽게 표현된 클러스터로 부터  
세부 전공 과목에 대한 정보를 제공



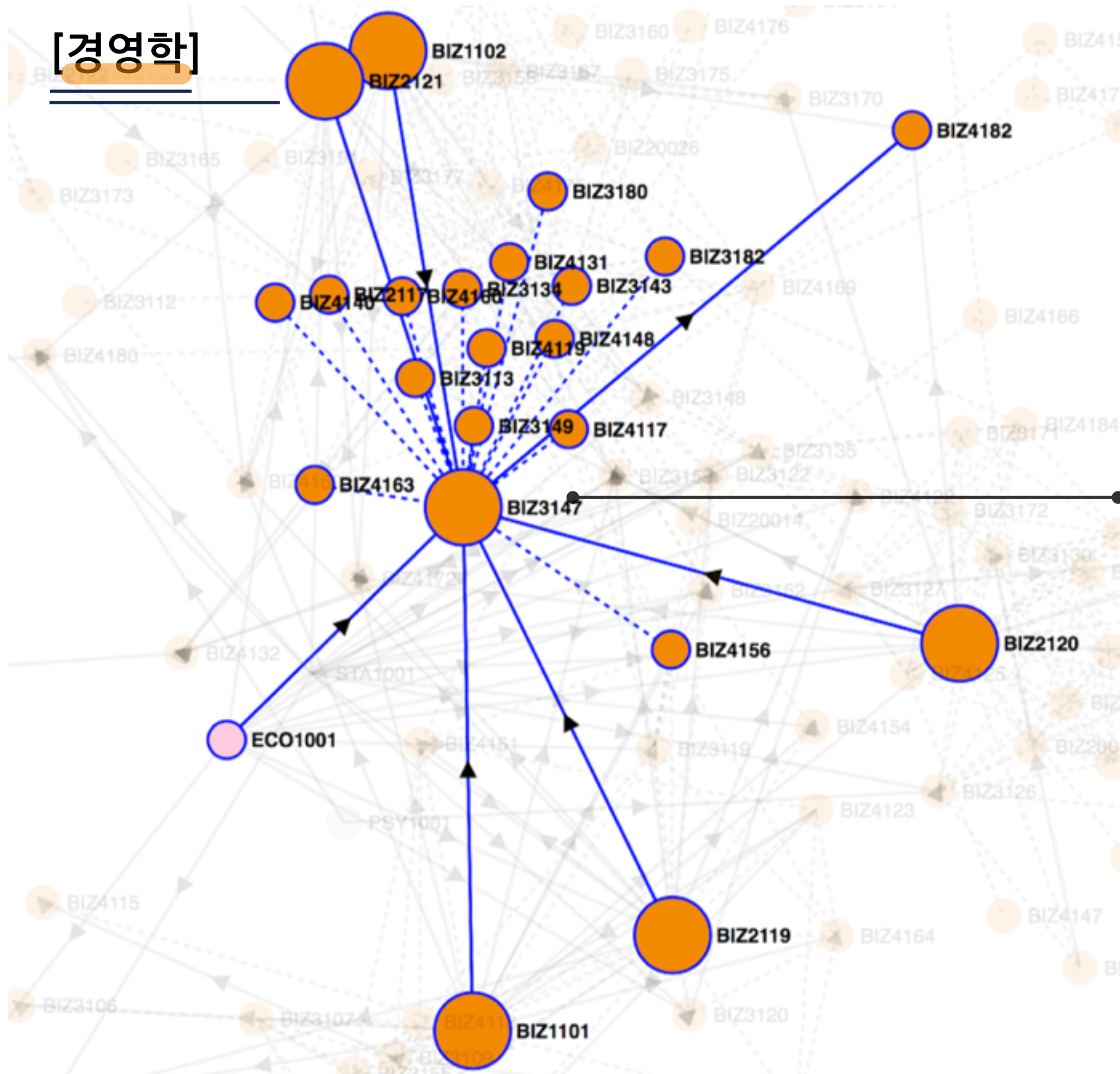


## [경영학]



- 기존 시스템에서는 어떤 과목을 수강한 이후의 연결 과목에 대한 정보를 제공하지 않았다.

## [경영학]



## “전략경영”은 어디쯤일까?

- 전략경영의 선수과목(안쪽을 향하는 화살)
- 전략경영의 다음과목(바깥을 향하는 화살)
- 내용이 유사한 과목(점선)

## 4-4. 학생의 융합적 학업 성취는 어떻게 표현할 수 있을까?

예시 [경영학 + 컴퓨터과학 이중전공]





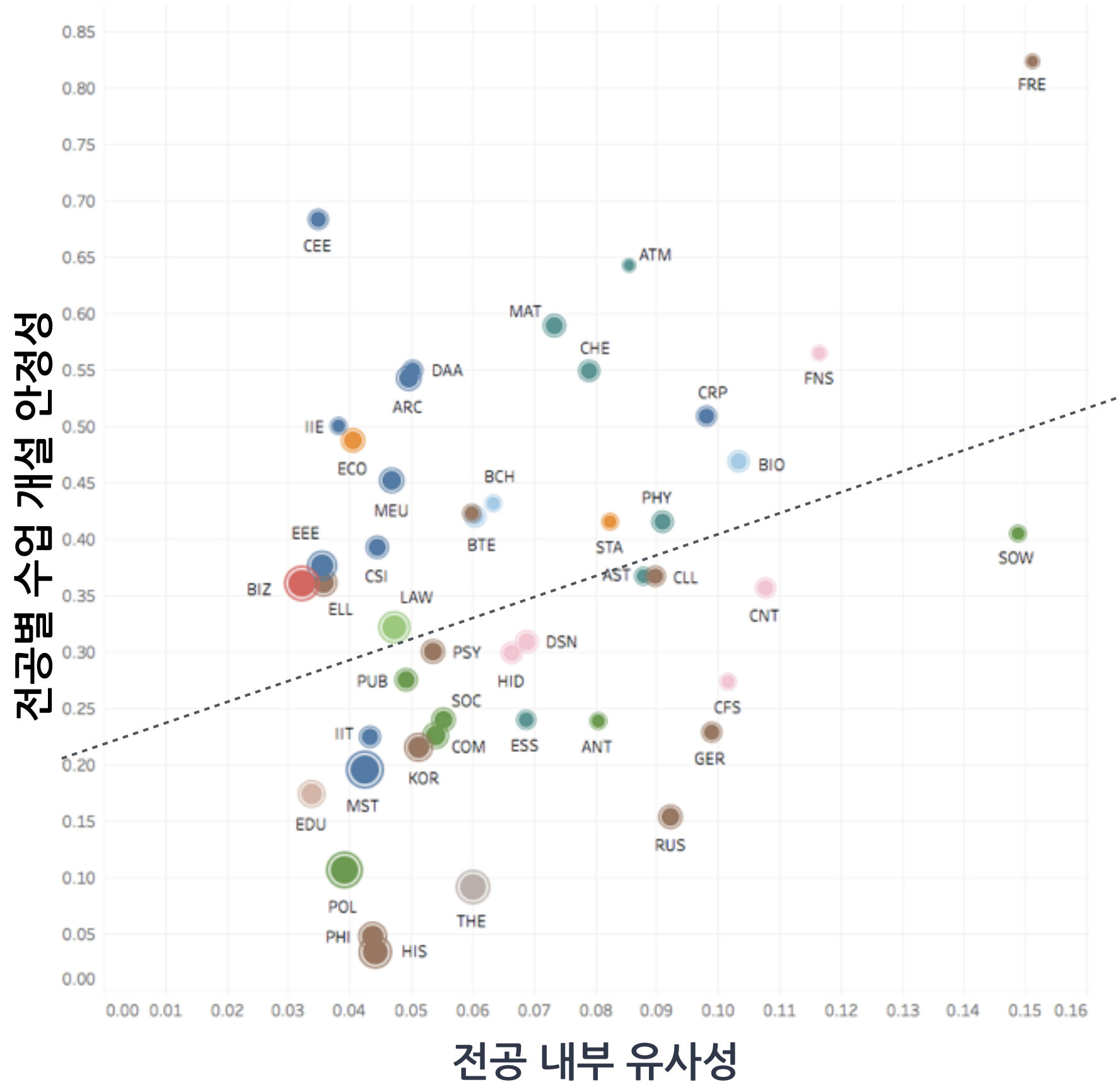
## 4-5. 전공별 커리큘럼 특성

### 1. 수업 개설 안정성

- 최소 3학기에 1회 이상  
개설되는 전공 과목의 비율

### 2. 전공 내부 유사성

- 전공 내부 과목 간 유사도의 평균값



## 4-5. 전공별 커리큘럼 특성

### 1. 수업 개설 안정성

- 최소 3학기에 1회 이상  
개설되는 전공 과목의 비율

### 2. 전공 내부 유사성

- 전공 내부 과목 간 유사도의 평균값

A)  유의미한 상관관계.



전공 내부 유사성

## 4-5. 전공별 커리큘럼 특성

### 1. 수업 개설 안정성

- 최소 3학기에 1회 이상  
개설되는 전공 과목의 비율

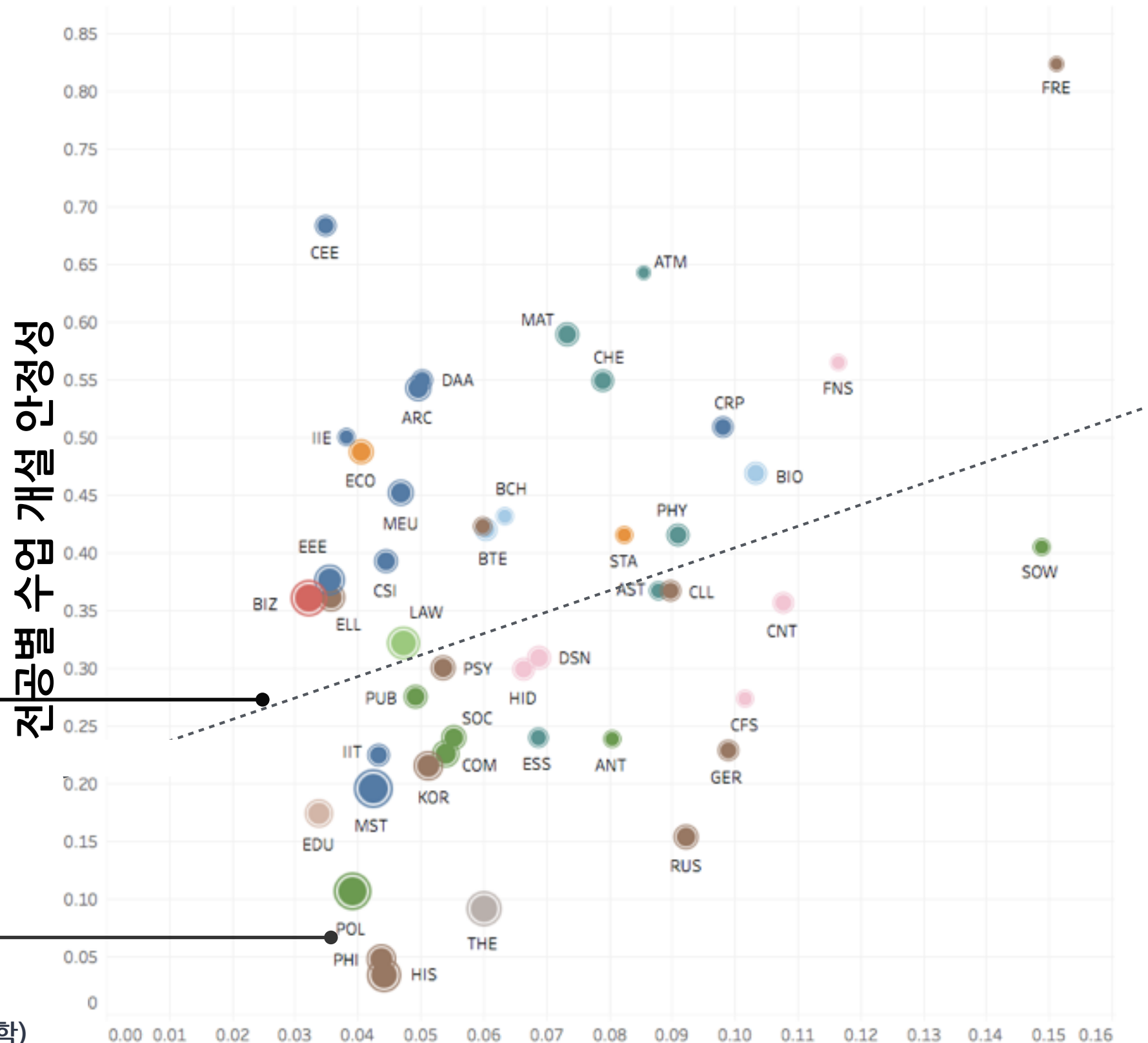
### 2. 전공 내부 유사성

- 전공 내부 과목 간 유사도의 평균값

A) 유의미한 상관관계.

B)

1. 시사 이슈에 민감한 연구 주제로 인해  
안정적으로 수업개설이 되지 않을 경우(정치학)
2. 교수진(faculty)의 크기가 작을 경우(사학, 철학)



전공 내부 유사성

## 4-5. 전공별 커리큘럼 특성

### 1. 수업 개설 안정성

- 최소 3학기에 1회 이상  
개설되는 전공 과목의 비율

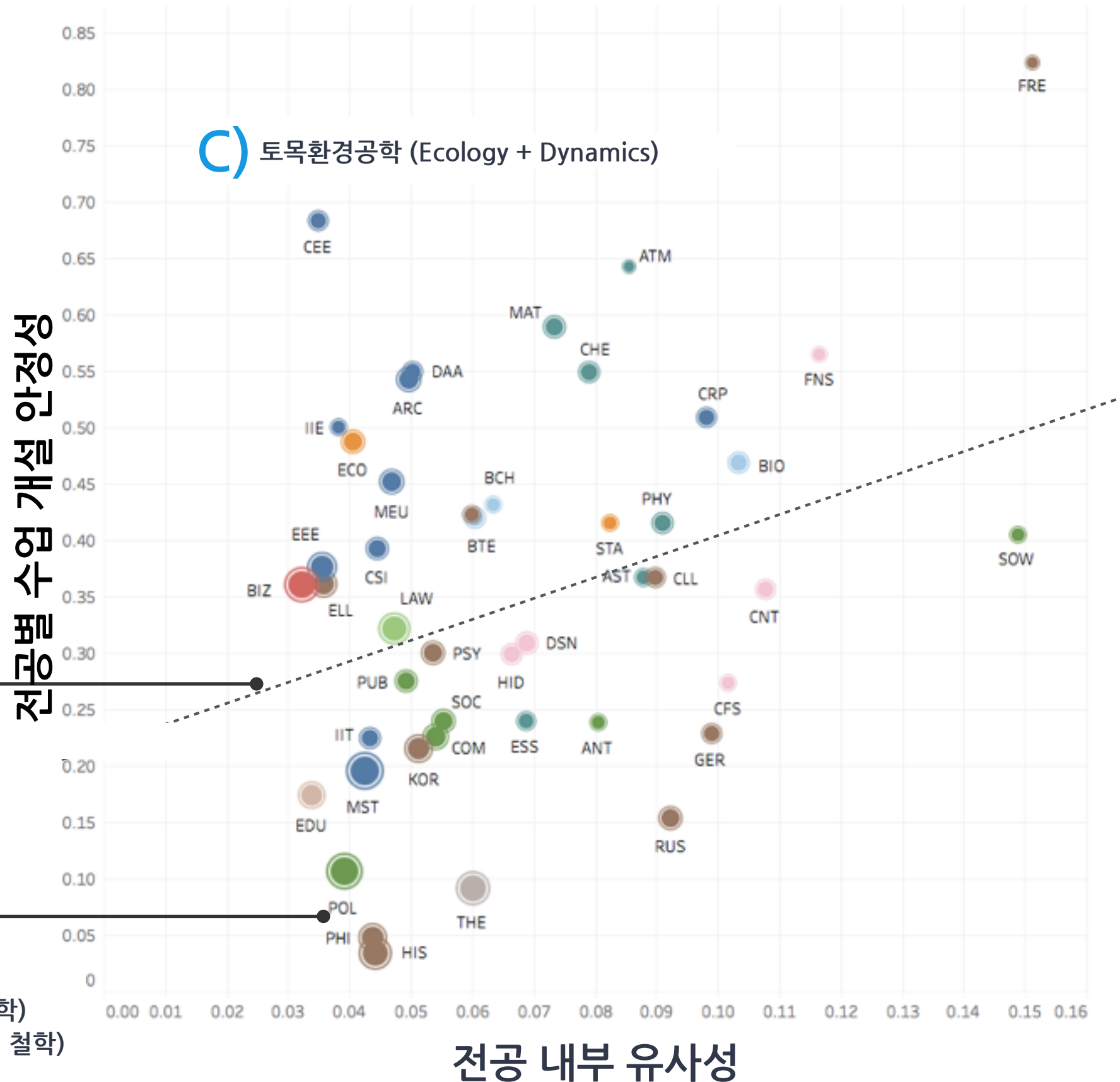
### 2. 전공 내부 유사성

- 전공 내부 과목 간 유사도의 평균값

A) 유의미한 상관관계.

B)

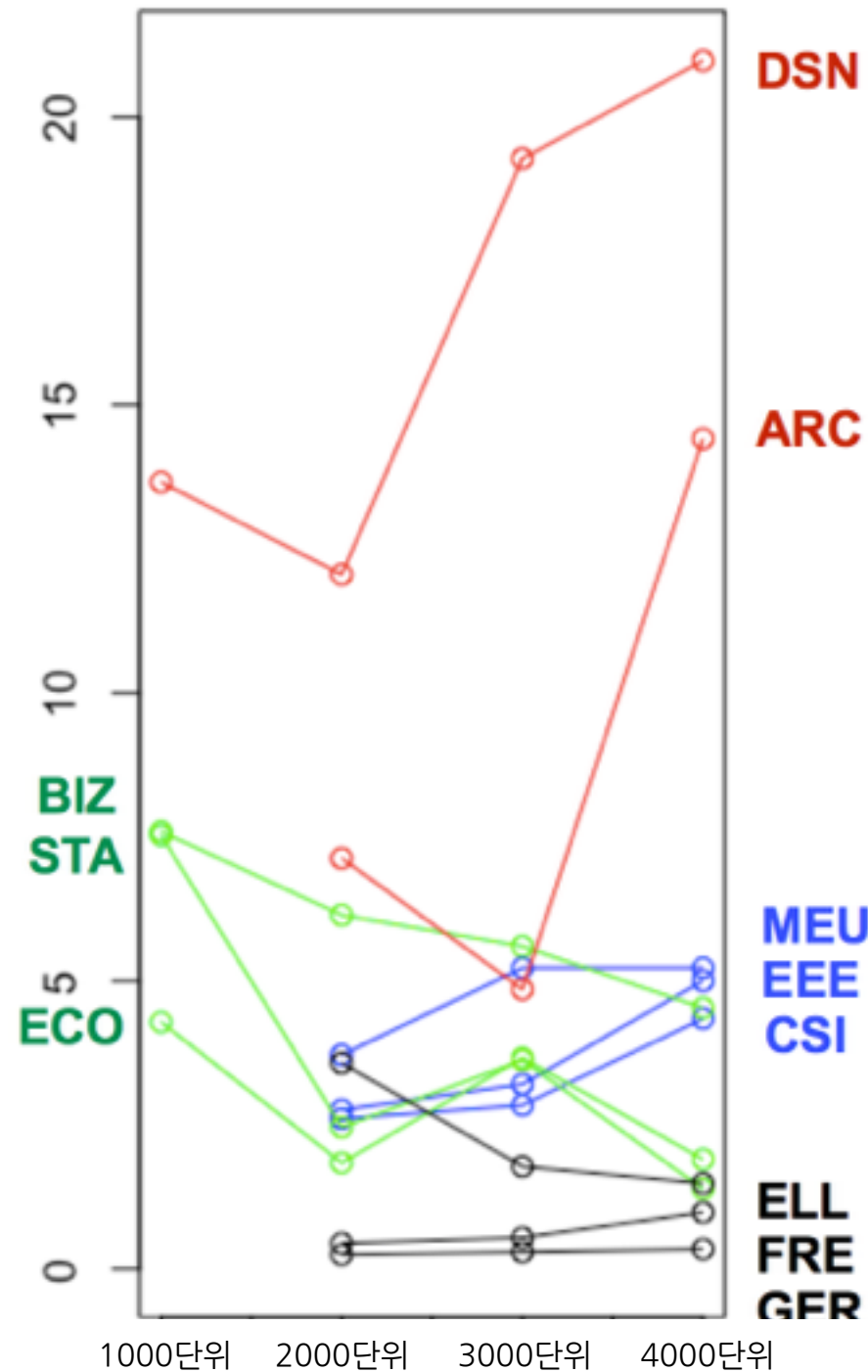
1. 시사 이슈에 민감한 연구 주제로 인해  
안정적으로 수업개설이 되지 않을 경우(정치학)
2. 교수진(faculty)의 크기가 작을 경우(사학, 철학)



## 4-5. 전공별 타학과 수업과의 유사도 분석

- (학정번호를 기준으로) 학년별 타학과 수업과의 유사도의 합
- 그 변화 추이에도 전공별 차이가 있음을 확인.

Business and Economics {



} Highly-integrated Majors ?

} Engineering Majors

} Language and Literature Majors



## 5. 연구 공헌

---

### 데이터마이닝 기법의 적용

- 네트워크 시각화, 분석 기법을 커리큘럼 분석에 적용.
- 강의 계획서(syllabus)에 대한 텍스트마이닝 연구는 그 동안 적극적으로 이루어지지 않았음.

### 데이터 중심의 커리큘럼 관리 시스템 개발

- 계량 분석을 통해 각 전공의 현황에 대한 진단을 시도.
- 복수의 대학에 대한 데이터가 확보 될 경우, 비교 분석을 통해, 대학별 세부 전공 특성화 현황과 커리큘럼 평가를 하는 것도 가능할 것으로 예상.

### 지식 연결 구조 분석

- 공동 학술 저자, 논문 키워드 간 연결 위주의 연구를 확장하여 지식의 부분집합(subset)이라고 할 수 있는 학부 수업에 네트워크 구조 이론을 적용하고 탐구.

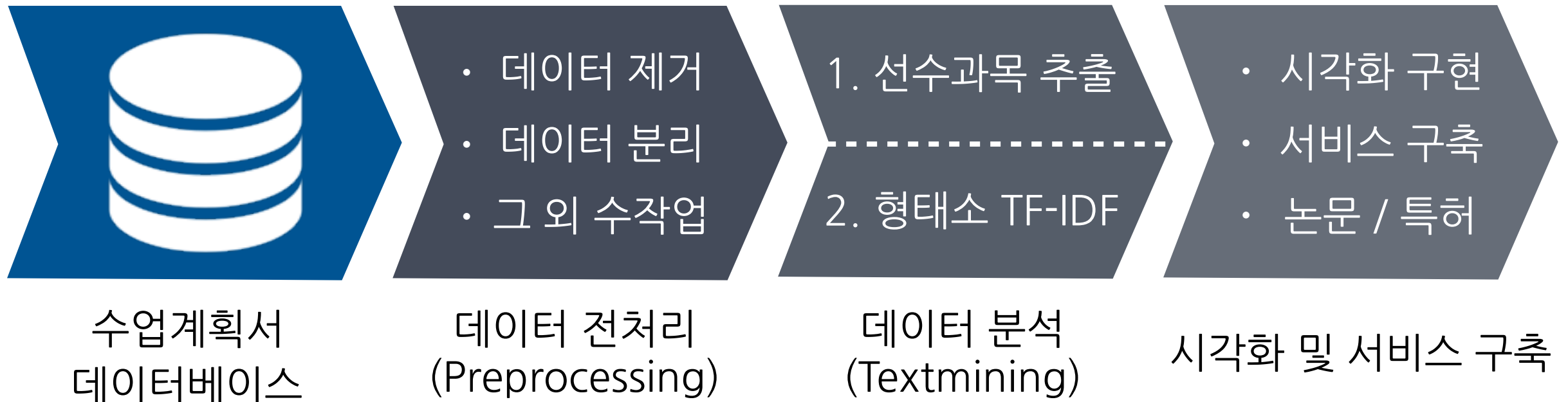
## 5. 연구 공헌 (Cont'd)

---

### 학생 중심적인 UI 개발

- 기존 수강편람 방식에서는 쉽게 제공할 수 없었던 강의간 연관 구조 / 선후관계 구조 등의 정보를 시각적으로 전달.
- 전공의 틀을 깬 정보 검색을 가능하게하여, 복수의 학술영역에 관심이 있는 학생들이 겪었던 정보 탐색의 어려움을 해소.
- 학생이 스스로 자신에게 맞는 비선형적인(Nonlinear) 학업 설계를 하는데 용이.
- 학생의 학업 성취에 대한 구체적인 정보를 제공.
- 인터랙티브 UI 예제와 베타 버전을 개발 중.

## 6. 진행 상황 및 후속 연구



### 진행 예정

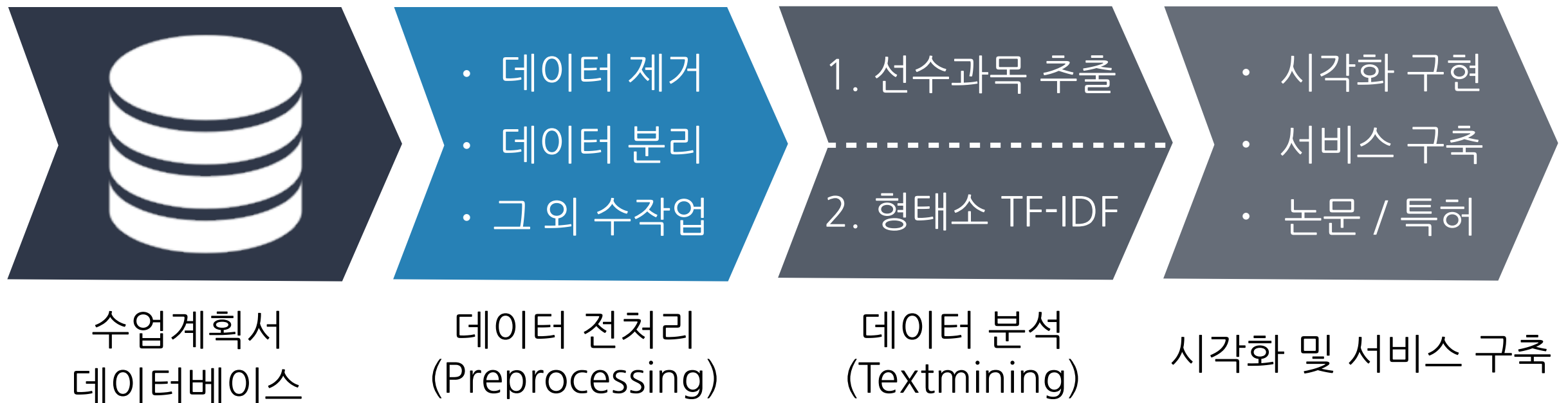
A. 2016년 봄학기부터 최신 데이터 확보 및 분석

### 후속 연구

A. 수업계획서 외 추가 데이터를 활용한 심층 연구

B. 타 대학의 수업계획서 데이터베이스를 활용한 비교 연구

## 6. 진행 상황 및 후속 연구



### 진행 예정

A. 수작업 데이터 처리

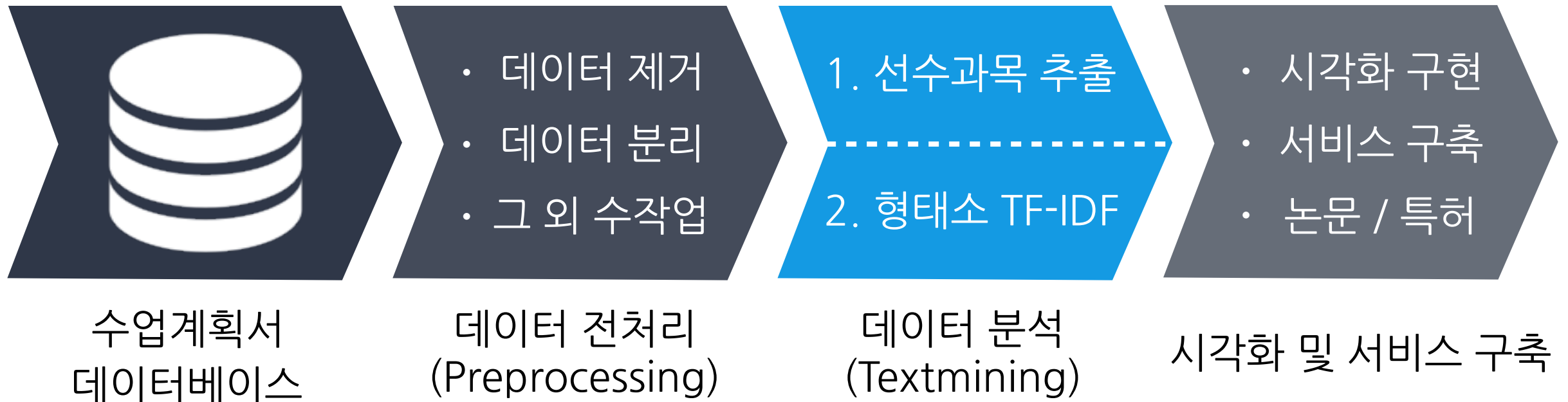
### 후속 연구

A. 텍스트마이닝 과정 개선에 대한 연구

- 번역, 다의어 어휘의 처리

B. UIC, 교양 과목, 처리가 어려웠던 과목(세미나, 합동 강의 등) 문제 해결

## 6. 진행 상황 및 후속 연구



### 진행 예정

A. 없음.

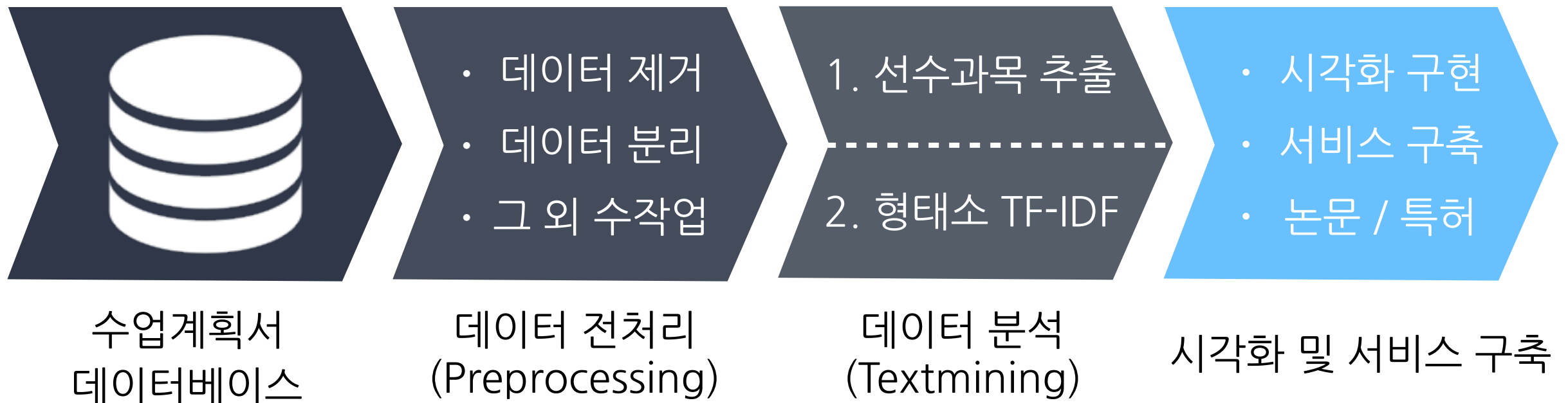
### 후속 연구

A. 전공별 현황에 대한 심층적 분석기법 연구  
- 교육학과와 협력

B. 유사도 분석에서는 드러나지 않는 연관 관계에 대한 연구 방법론 개발



## 6. 진행 상황 및 후속 연구



### 진행 예정

- A. 논문화 작업(진행률 70%)
- B. 특허 출원 예정(진행률 40%)
- C. 인터랙티브 검색 서비스(Beta) 개발(진행률 20%)
- D. 개발 후 UX의 효과적인 개선 여부 확인을 위한 설문 및 평가 예정

### 후속 연구

- A. 많음.