

# ‘미래 영향 환경변수 및 시나리오 도출 연구’ 기술자문 결과

2018년 10월

작성자 : KISTI 미래기술분석센터 박진서, 이준영



# 차례

---

- 1 데이터 수집 및 1차 분석
- 2 저자 및 인덱스 키워드 분석 (2차 분석)
- 3 Overlay Mapping 분석 (3차 분석)
- 4 미디어 분석 (4차 분석)

## 데이터 수집 및 1차 분석

## 데이터 수집 및 1차 분석

### ▪ 데이터셋 확보 및 최종 분석데이터 확정

#### 1) 1차 검색

- KISTI가 보유한 SCOPUS DB 대상
- 검색식: TS=(("futures" OR "foresight\*") OR ("future\*" AND "prediction\*")) PY=(2008-2017)
- 검색결과: 68,642건

#### 2) 관련성 없는 데이터 제거

- 6명의 미래연구원 소속 연구원이 68,642건에 대해 논문 제목 및 요약 확인 과정을 거쳐 미래연구와 관련 없는 12,483건 데이터 제외
- 분석대상: 56,159건

### ▪ 키워드 선택 범위 설정

#### 1) 필드별 데이터 커버리지

- Author Keyword: 74%
- Index Keyword: 82%

#### 2) 키워드 개수

- Author Keyword: 101,885개(정제전)
- Index Keyword: 158,464개(정제전)

#### 3) Author Keyword와 Index Keyword를 하나의 필드로 통합할 경우,

- 218,694개 키워드(정제전)
- 커버리지는 99%



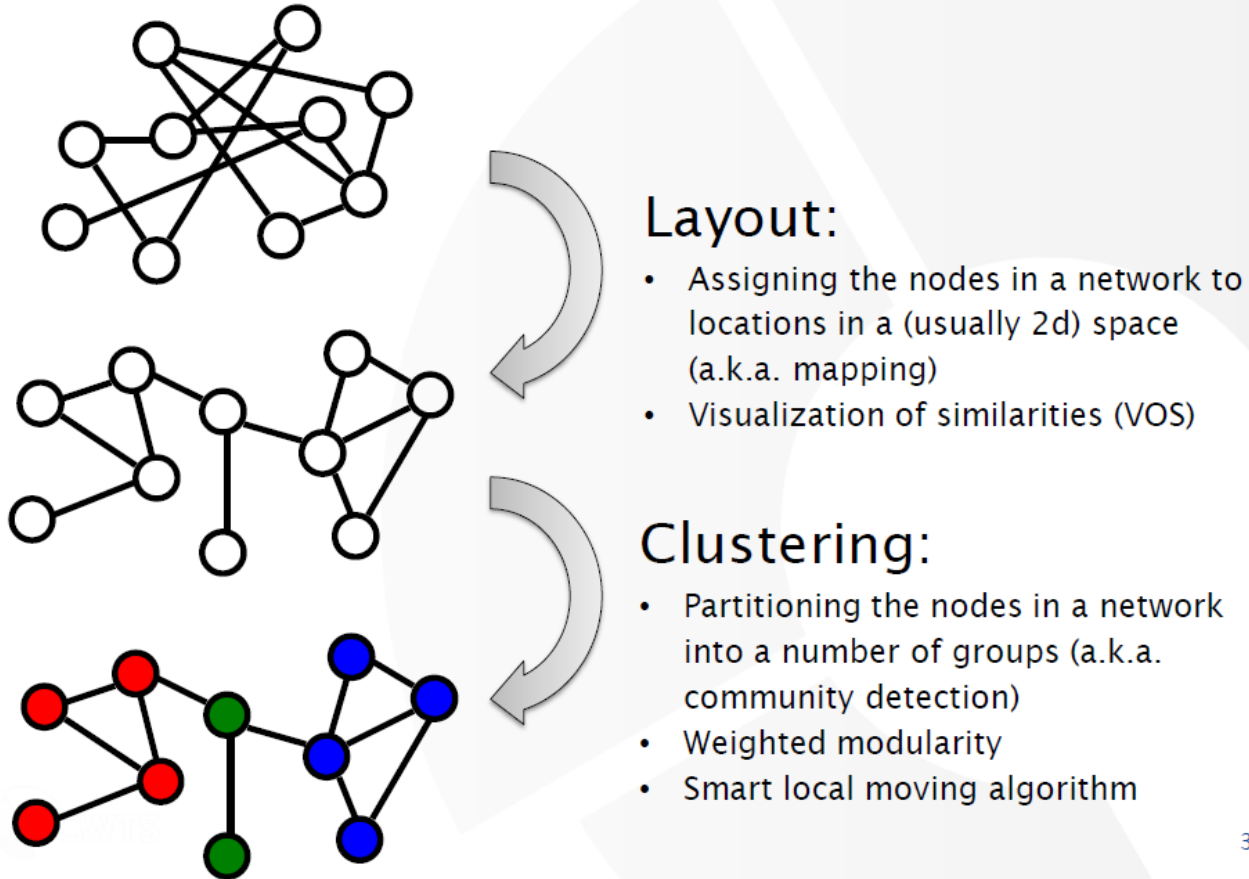
## 데이터 수집 및 1차 분석

- 일반적으로 단어들 간의 관계를 분석하는 방법은
    - 1) co-word mapping based on social network analysis,
    - 2) co-word mapping based on latent semantic analysis,
    - 3) topic modelling 등이 있음.(Leydesdorff, 2014; Leydesdorff and Nerghe, 2016)
  - 문서의 속성, 단어의 추출 방법 등에 따라 어느 방법이 최선인지는 아직 알려지지 않았음.
  - **본 자문에서는 가장 직관적인 1) co-word mapping based on social network analysis를 사용**
  - 즉, 논문과 특허에 등장하는 키워드가 하나의 문서에서 얼마나 동시출현하였는가를 기준으로 동시출현행렬 생성, 동시출현행렬 자료를 기초로 클러스터링 및 맵핑 수행
- 
- 본 연구의 목적은 분석대상인 지식맵을 구성하는 연구주제가 무엇인지 확인(clustering)하고, 연구주제 간의 관계를 시각화(mapping)하는 것임.
  - 맵핑 및 클러스터링은 VOSviewer 알고리즘으로 수행
  - VOSviewer에서 두 단어 간의 유사도는 두 단어가 한 논문에서 동시출현한 횟수에 따라 비례하고, 두 단어의 다른 단어들과 동시 출현한 빈도의 합에 반비례
  - VOSviewer는 맵핑과 클러스터링을 동시에 수행
    - 먼저 2차원 공간에서 개별 단어의 좌표는 모든 두 단어들 간의 거리와 두 단어의 유사도를 가중치로 곱한 합이 최소화되게 설정
    - 즉, VOSviewer 맵핑에서 서로 높은 유사도를 갖는 단어들은 가깝게 위치하고, 유사도가 낮으면 두 단어는 멀리 위치
  - VOSviewer에서 클러스터링은 weighted network에서 주로 사용하는 modularity-based clustering의 일종

## 데이터 수집 및 1차 분석

- VOSviewer의 클러스터링 및 맵핑 과정

### Network analysis techniques



34

## 데이터 수집 및 1차 분석

- VOSviewer의 클러스터링 및 맵핑 과정

# Unified approach to mapping and clustering

$$\text{Minimize } Q(x_1, \dots, x_n) = \sum_{i < j} \frac{2m}{k_i k_j} A_{ij} d_{ij}^2 - \sum_{i < j} d_{ij}$$

where

$n$ : number of nodes in the network

$m$ : total weight of all edges in the network

$A_{ij}$ : weight of edge between nodes  $i$  and  $j$

$k_i$ : total weight of all edges of node  $i$

### Clustering

$x_i$ : integer denoting the community to which node  $i$  belongs

$$d_{ij} = \begin{cases} 0 & \text{if } x_i = x_j \\ 1/\gamma & \text{if } x_i \neq x_j \end{cases}$$

$\gamma$ : resolution parameter

### Mapping

$x_i$ : vector denoting the location of node  $i$  in a  $p$ -dimensional space

$$d_{ij} = \|x_i - x_j\| = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2}$$

\* 출처: van Eck & Waltman(2017). Science Mapping and Research Positioning, 2017 BenchTech Seminar, Technical University Munich, Munich, Germany, June 28, 2017.

## 데이터 수집 및 1차 분석

### ■ 키워드맵의 의미와 전제

- Science maps are **symbolic representations** of scientific fields or organizations in which the elements of the map are associated with topics or themes. Elements
- Maps are built **on the basis of a matrix of similarity measures** computed from correlation functions among information items present in different elements (e.g., cooccurrence of the same author in various articles).

### ■ 맵핑 분석의 장점

- ① Maps position units in a (two-dimensional [2D]) network instead of ranking them on a (one-dimensional) list. As in any data visualization technique, maps furthermore **facilitate the reading of bibliometric information by nonexperts** — with the downside that they also leave room for manipulating the interpretation of data structures.
- ② Second, maps allow **for the representation of diverse and large sets of data** in a succinct way.
- ③ Third, precisely because they make it possible to combine different types of data, maps also enable users to **explore different views on a given issue. This interpretive flexibility induces reflexive awareness** about the phenomenon the user is analysing and about the analytical value (and pitfalls) of these tools. Implicitly, science maps convey a key message: bibliometrics cannot provide definite, “closed” answers to science policy questions, such as “picking the winners.” Instead, maps remain more explicitly heuristic tools to explore and potentially open up plural perspectives to inform decisions and evaluations

\* 출처: Rafols, Porter and Leydesdorff(2010), Science Overlay Maps: A New Tool for Research Policy and Library Management, Journal of the American Society for Information Science and Technology, Vol. 61, No. 9, pp. 1871–1887.

## 데이터 수집 및 1차 분석

- VOSviewer를 통한 가시화
- 노드 간의 관계를 보여 주는 일반적인 network 가시화(network view)뿐 아니라 VOSviewer에서는 density 가시화(density view) 기능도 제시
  - 특정 단어의 밀도는 이웃하는 단어의 수와 이웃하는 아이템의 중요성을 반영, 따라서 density view는 전체적인 맵의 구조를 개괄하는데 유용하며 맵에서 어느 영역이 가장 중요한 영역인지 판단할 수 있음.
  - network view에서 통계적 클러스터링 알고리즘에 의한 클러스터의 분포를 확인할 수 있다면, density view에서는 전체 맵에서 중요 영역을 직관적으로 확인할 수 있다는 장점이 있음.
  - 즉, 통계적 클러스터링 알고리즘은 중요하지 않은 단어들도 포함하여 sub topic을 구분하여 보여 준다면, density view에서는 직관적으로 '중요한 단어'들의 집합을 확인가능하다는 장점이 있음

### ▪ 관련 참고문헌

van Eck, N. and Waltman, L.(2010), Software survey: VOSviewer, a computer program for bibliometric mapping, *Scientometrics*, Vol. 84, pp. 523-538.

van Eck, N. J., Waltman, L., Dekker, R. and van den Berg, J.(2010), A Comparison of Two Techniques for Bibliometric Mapping: Multidimensional Scaling and VOS, *Journal of the American Society for Information Science and Technology*, Vol. 61, No. 12, pp. 2405-2416.

Waltman, L., van Eck, N. J. and Noyons, E. C. M.(2010), A unified approach to mapping and clustering of bibliometric networks, *Journal of Informetrics*, Vol. 4, pp. 629-635.

Leydesdorff, L.(2014), Science Visualization and Discursive Knowledge, in Cronin, B. and Sugimoto, C. R. (eds.), *Beyond Bibliometrics – Harnessing Multidimensional Indicators of Scholarly Impact*, MIT Press.

Leydesdorff, L. and Nerghes, A.(2016), Co-word and Topic Modelling – a comparison using small and medium-sized corpora (n<1000), *Journal of the Association for Information Science and Technology (JASIST)*, in press







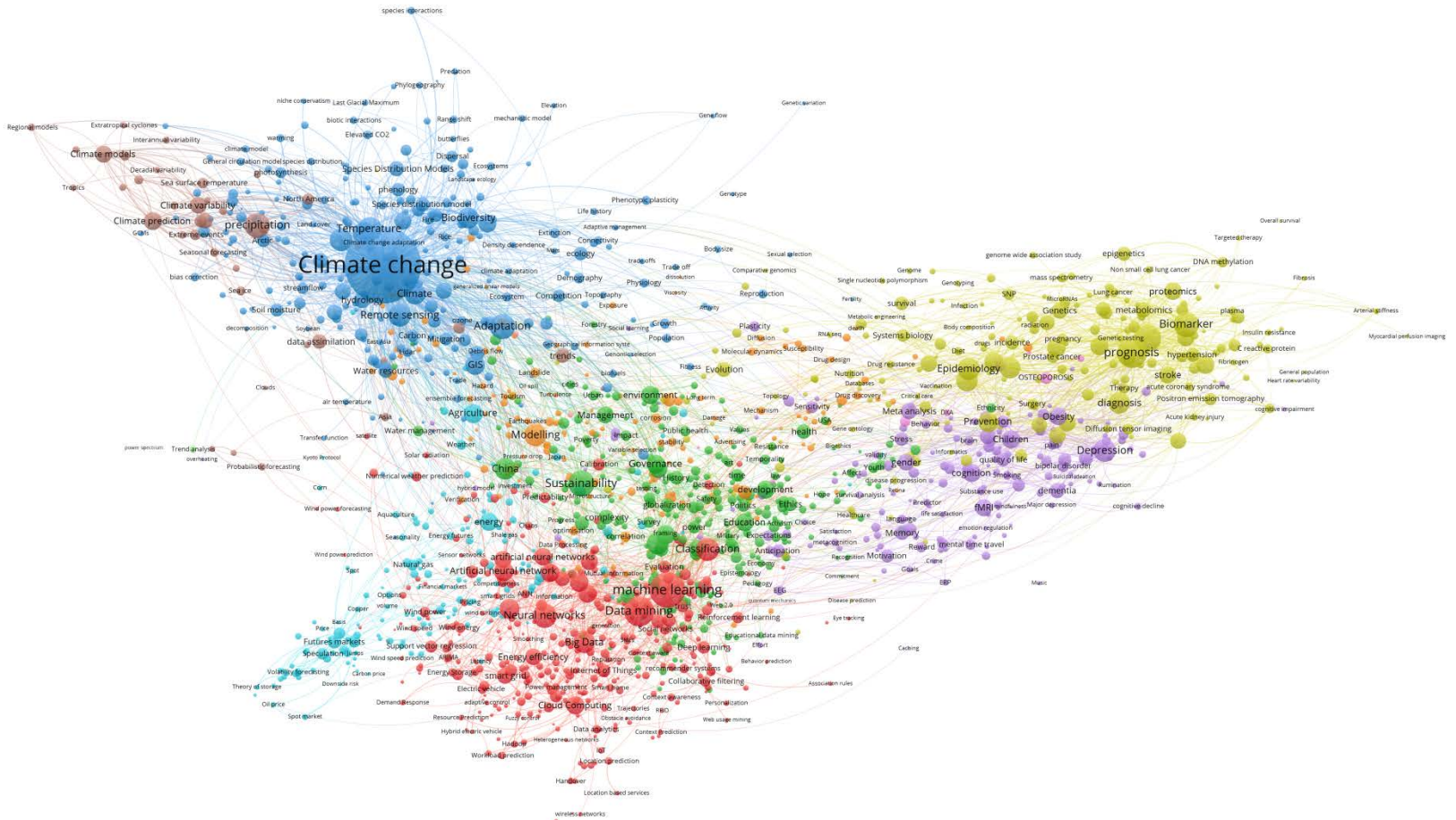




저자 및 인덱스 키워드 분석 (2차 분석)

## 저자 및 인덱스 키워드 분석 (2차 분석)

### 저자키워드: 네트워크맵











## Overlay Mapping 분석 (3차 분석)

## Overlay Mapping 분석 (3차 분석)

### Overlay Mapping의 장점

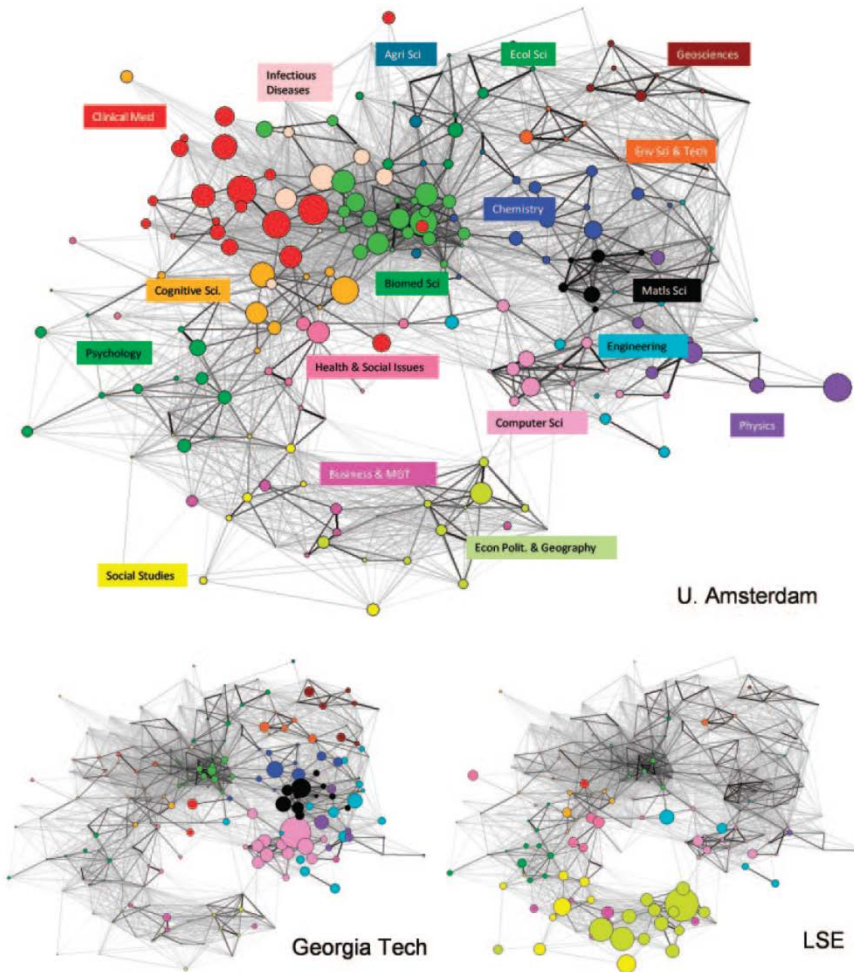


FIG. 3. Publications profiles of the University of Amsterdam, Georgia Tech, and London School of Economics (LSE) overlaid on the map of science.

- First, they provide a **visual framework** that enables us to make immediate and **intuitively rich comparisons**.
- Second, they use **cognitive units** for the representation (disciplines and specialties) that fit with conventional wisdom, whereas one can expect the analytical aggregates of local maps to be unstable and difficult to interpret.
- Third, whereas the generation of meaningful local maps requires bibliometric expertise, overlay maps can be produced SCI users, **who are not experts in scientometrics**.
- Finally, they **can be used for various purposes** depending **on the units of analysis displayed by the size of the nodes**, whether number of publications, citing articles, cited references, growth or other indicators, as shown by a series of recent studies.

\* 출처: Rafols, Porter and Leydesdorff(2010), Science Overlay Maps: A New Tool for Research Policy and Library Management, Journal of the American Society for Information Science and Technology, Vol. 61, No. 9, pp. 1871-1887.























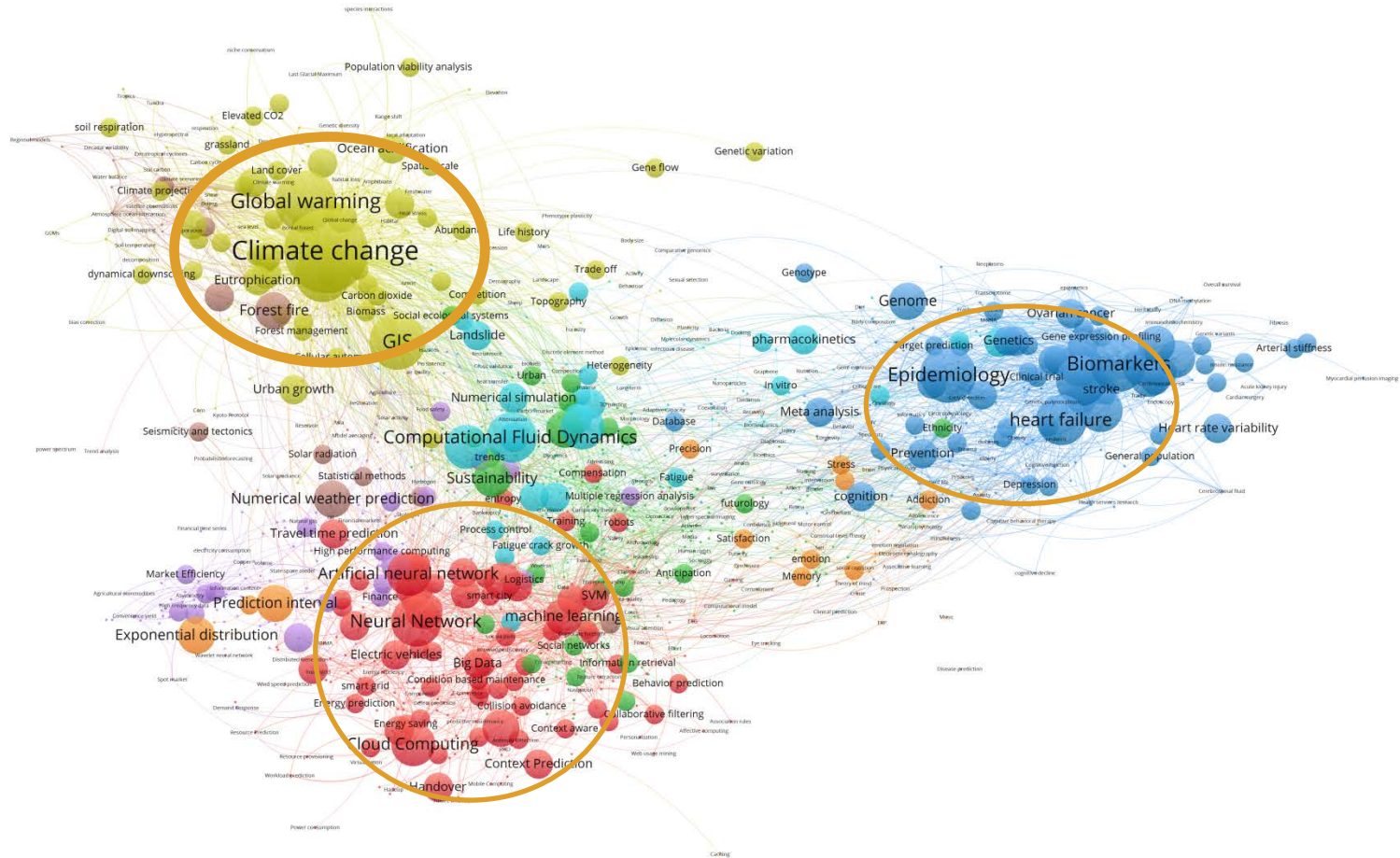






## Overlay Mapping 분석 (3차 분석)

- 저자키워드: 네트워크맵 (일본)













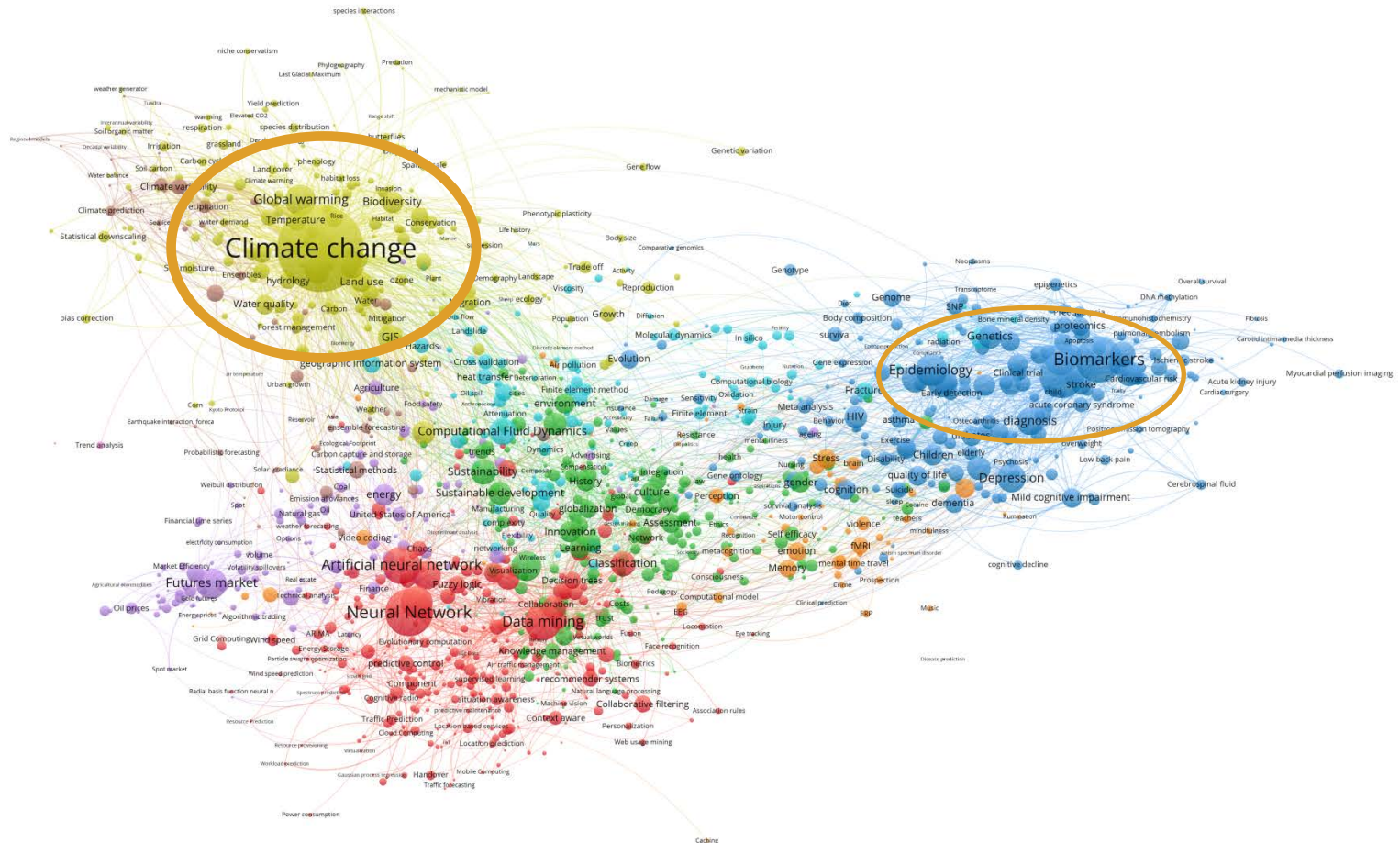






## Overlay Mapping 분석 (3차 분석)

- 저자키워드: 네트워크맵 (2008~2009)













































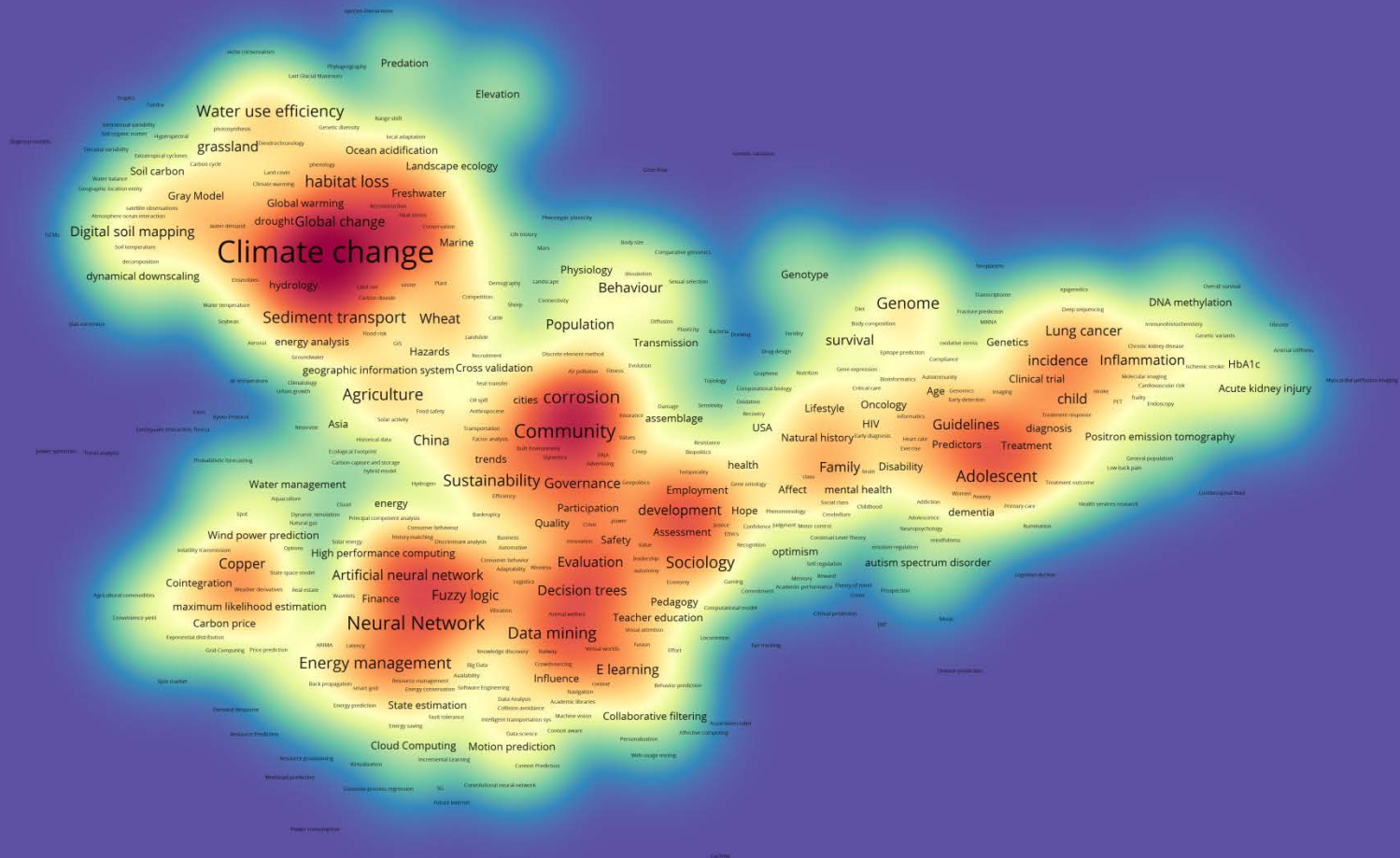






# Overlay Mapping 분석 (3차 분석)

- 저자키워드: 밀도맵 (호주)





















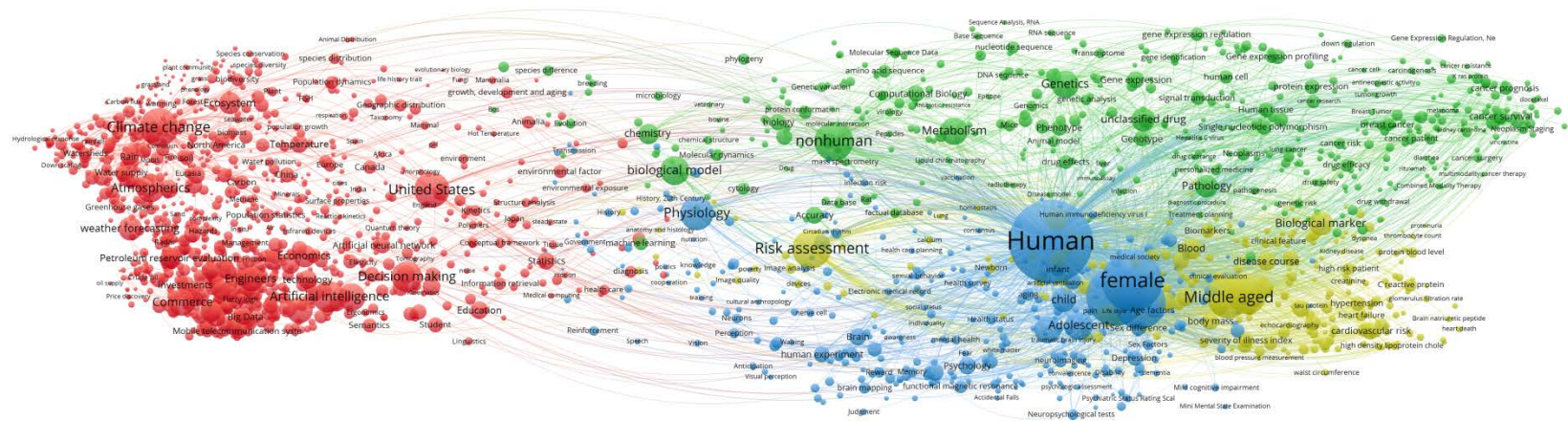






## Overlay Mapping 분석 (3차 분석)

- 인덱스키워드: 네트워크맵 (미국)





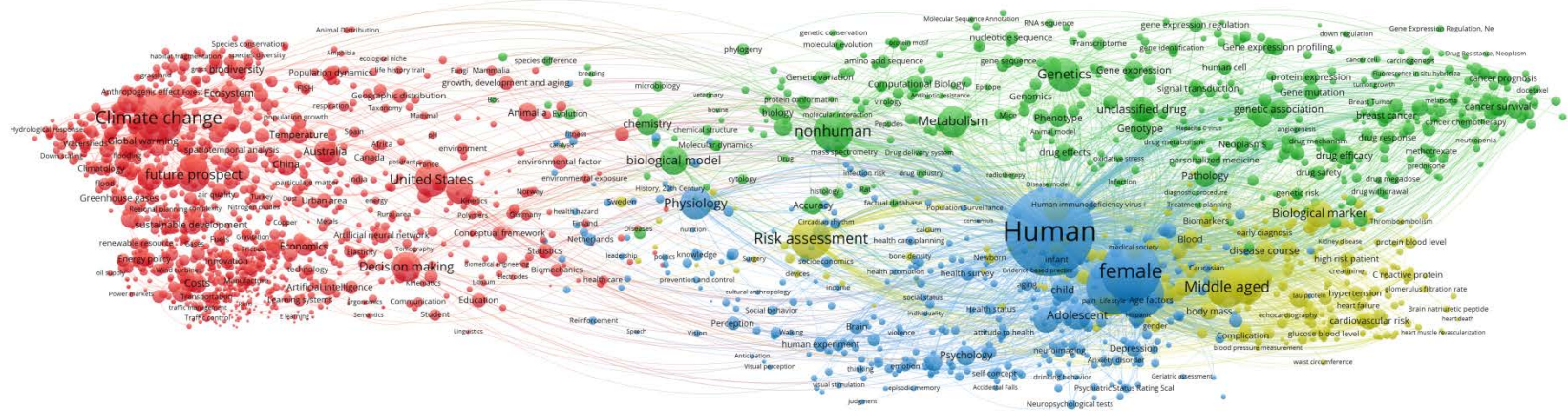






## Overlay Mapping 분석 (3차 분석)

- 인덱스키워드: 네트워크맵 (영국)



















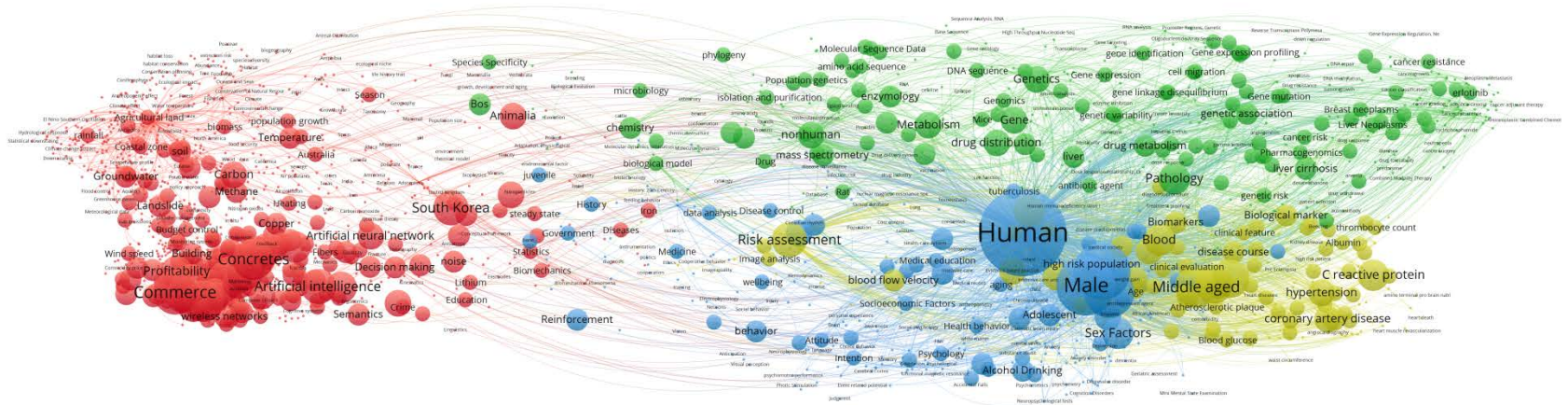






## Overlay Mapping 분석 (3차 분석)

- 인덱스키워드: 네트워크맵 (한국)



## Overlay Mapping 분석 (3차 분석)

- 인덱스키워드: 네트워크맵 (2008~2009)

