
공공데이터에 적합한 다양한 소셜 그래프 비주얼라이제이션 알고리즘 제안

Social graph visualization techniques for public data

이만재, Manjai Lee*, 온병원, Byung-Won On**

요약 최근 다양한 공공데이터가 개방되고 있으며, 적절한 데이터 마이닝과 시각화 알고리즘을 통해 일반 시민에게 서비스 되고 있다. 이를 통해 정부와 지방자치단체는 공공 정책의 투명성과 효율성을 널리 알릴 수 있을 뿐 아니라, 일반 사용자들이 개방된 공공데이터를 재가공하여 서비스함으로써 관련 산업의 성장을 이끌고 있다. 공공데이터의 최종 사용자는 일반 시민이기 때문에, 누구나 손쉽게 이해할 수 있도록 공공데이터를 적절히 시각화하는 것이 무엇보다 중요하다. 본 연구에서는 공공데이터 비주얼라이제이션의 중요성을 널리 알리기 위해, 일반 국민이 관심을 가질만한 공공데이터로 UN 회원국의 투표 데이터를 고려한다. 외교와 교육 목적으로 그 활용 가치가 높고 데이터를 쉽게 얻을 수 있는 장점이 있다. 또한 적절한 데이터 마이닝과 시각화 과정을 거친다면, 일반 사용자들이 유엔 회원국 간의 투표 성향에 대한 통찰력을 쉽게 얻을 수 있다. 유엔 투표 데이터를 시각화하기 위해서는, 회원국 간의 투표성향 유사도를 측정하고, 이를 바탕으로 소셜 그래프를 구현한다. 그리고 그래프 레이아웃 알고리즘을 적용하여 그래프를 화면에 렌더링 하게 된다. 기존 방법을 이용하여 소셜 그래프를 비주얼라이제이션 할 경우에 그래프의 복잡도가 증가하여 유엔 회원국 간의 투표성향을 파악하는데 큰 어려움이 있다. 이러한 문제를 개선하기 위해, 본 논문에서는 친구 매칭(Friend-Matching), 친구-라이벌 매칭(Friend-Rival Matching), 버블힙(Bubble Heap) 알고리즘들을 차례로 제안한다. 제안된 알고리즘을 바탕으로, 기존 그래프 비주얼라이제이션을 개선하여 일반 사용자들이 손쉽게 유엔 회원국 간의 투표성향과 관련된 특정 패턴이나 통찰력을 얻는데 큰 도움을 줄 것이다. 또한 웹에서 동작하는 프로토타입을 구현하여, 누구나 방문하여 테스트를 할 수 있다. 웹 페이지 주소: <http://datalab.kunsan.ac.kr/politiz/un/>

Abstract Nowadays various public data have been serviced to the public. Through the opening of public data, the transparency and effectiveness of public policy developed by governments are increased and users can lead to the growth of industry related to public data. Since end-users of using public data are citizens, it is very important for everyone to figure out the meaning of public data using proper visualization techniques. In this work, to indicate the significance of widespread public data, we consider UN voting record as public data in which many people may be interested. In general, it has high utilization value by diplomatic and educational purposes, and is available in public. If we use proper data mining and visualization algorithms, we can get an insight regarding the voting patterns of UN members. To visualize, it is necessary to measure the voting similarity values among UN members and then a social graph is created by the similarity values. Next, using a graph layout algorithm, the social graph is rendered on the screen. If we use the existing method for visualizing the social graph, it is hard to understand the meaning of the social graph because the graph is usually dense. To improve the weak point of the existing social graph visualization, we propose Friend-Matching, Friend-Rival Matching, and Bubble Heap algorithms in this paper. We also validate that our proposed algorithms can improve the quality of visualizing social graphs displayed by the existing method. Finally, our prototype system has been released in <http://datalab.kunsan.ac.kr/politiz/un/>. Please, see if it is useful in the aspect of public data utilization.

핵심어 : 공공데이터, 소셜 그래프, 비주얼라이제이션, 유사도, 그래프 레이아웃 알고리즘
Public data, social graph, visualization, similarity, graph layout algorithms

본 논문은 2012년 차세대융합기술연구원 학술 연구비 지원에 의하여 연구되었음. (No. 2012-P3-22)

*주저자 : 차세대융합기술연구원 특임연구위원

**교신저자 : 군산대학교 통계컴퓨터과학과 조교수: e-mail : bwon@kunsan.ac.kr

■ 접수일 : 2014년 8월 19일 / 심사일 : 2014년 9월 27일 / 게재확정일 : 2015년 1월 15일

1. 서론

공공데이터는 정부 또는 공공기관이 보유하고 있는 데이터를 말한다. 공공기관의 업무과정에서 얻어진 데이터들로 활용 가치가 높고, 일반에게 공개될 경우 정부의 효율적인 공공 정책을 널리 알릴 수 있다. 또한 관련 산업의 성장을 유도할 수 있는 장점이 있다. 최근에 선진 각국에서는 공공데이터 활용이 정부의 대국민 서비스와 직결되어 있음을 인지하고, 공공데이터 활용을 정부 정책으로 적극 추진하고 있다. 2009년 미국의 오바마 대통령은 data.gov 웹사이트를 구축하고, 정부 데이터를 공개토록 하였으며, 영국 정부는 웹의 창시자인 팀 버너스 리를 책임자로 기용하여 공공데이터를 링크드 데이터 방식으로 제공하도록 하였다. 이와 같이 일반 국민이 공공데이터를 활용할 수 있도록 공공데이터의 개방 및 인프라를 구축하는데 많은 노력을 기울이고 있다. 반면, 예산, 복지, 환경 등 각각의 공공데이터는 데이터 포맷과 특성이 상이하기 때문에, 개별 공공데이터의 특성을 파악한 후에 적절한 데이터 마이닝 기법을 사용하여 통찰력을 얻고, 이를 시각화하여 일반 국민에게 공공 정책을 쉽게 알리는 것이 필요하다. 특히 공공데이터의 의미를 일반 시민에게 쉽게 알리기 위해서는 비주얼라이제이션을 사용해야 한다. 숫자 데이터의 경우 전통적인 막대그래프나 원 그래프를 사용하며, 텍스트의 빈도를 나타내거나 할 경우에는 워드 트리나 태그 클라우드를 사용한다. 또한 사물 간의 관계를 표시하고자 할 경우에는 소셜 그래프로 나타내고 사물 간의 계층적인 시각화를 위해서는 트리맵이 사용된다[7].

이를 잘 보여주는 예로, 유럽의 에너지라는 비주얼라이제이션이 있다. 이것은 유럽 국가의 에너지 생산, 이산화탄소 배출, 재생 에너지 등 에너지 정책과 관련된 데이터를 애니메이션을 통해 보여준다[13]. 영국의 경우에는 국민이 낸 세금이 어떻게 사용되는지를 예산항목과 함께 지역마다 어떻게 다르게 사용되는 가를 보여준다[14]. OECD는 동일한 기준으로 회원국 간의 통계를 비교할 수 있도록 다양한 데이터를 생산하여 제공한다. 주택, 수입, 직업, 교육 등에서 회원국의 삶의 질을 쉽게 비교할 수 있도록 시각화 하였다[12].

본 연구에서는 공공데이터 비주얼라이제이션의 중요성을 널리 알리기 위해, 일반 국민이 관심을 가질만한 공공데이터로 UN 회원국의 투표 데이터를 고려한다. 외교와 교육 목적으로 활용 가치가 클 뿐 아니라 데이터[15]를 쉽게 얻을 수 있는 장점이 있다. 또한 적절한 마이닝과 시각화를 거친다면, 일반 사용자들도 쉽게 통찰력을 얻을 수 있다.

UN은 전 세계적으로 가장 중요한 국제기구이다. 1946년에 51개 회원국으로 출발하였으며 현재 193개 회원국이 가입되어 있다. UN총회는 회원국의 투표로 중요한 안건을 결정한다. 국제기구인 만큼 결정된 내용의 문서화는 중요한 문제이며 설립부터 지금까지의 모든 결의안에 대한 기록이 유지되고 있다. 이중 본회의 투표는 전 세계의 국가 간의 이해관계를 파악하고자

할 경우 중요한 자료가 된다. 그러나 적당한 분석도구가 없다면 UN 총회에서 국가 간의 변화를 파악하는 것은 개인의 노력으로는 불가능하다. 이를 테면 UN 투표 데이터는 1946년부터 2012년 말까지 총 5,211건을 포함한다. 만약 전체 안건을 모두 조사, 분석하여 UN 회원국 간의 투표성향을 알고자 한다면, 많은 연구자들이 많은 시간을 투자하여 그 의미를 파악하여야 한다. 하지만 회원국 간의 투표성향을 나타내는 소셜 그래프를 구현하여 웹에서 보여주고, 사용자의 인터랙션에 반응하여 적절한 그래프를 보여준다면, 일반 사용자들이 손쉽게 유엔 투표 데이터로부터 새로운 의미를 파악할 수 있다. 또한 다른 공공데이터와 달리, UN의 분석은 다음과 같이 해결하기 어려운 새로운 문제가 발생한다. 첫째 분석대상이 1946년부터 2012년까지 67년의 기간으로 하나의 그래프가 아니라 연속적인 그래프로 구현하여야 한다. 둘째, 회원국 수의 변화에 따른 노드 수의 변화를 수용할 수 있어야 한다. 마지막으로 회원국의 경제적 능력, 인구 등 데이터를 사용해서 이해하기 쉬운 비주얼라이제이션을 구현하는 문제가 뒤따른다.

본 논문에서는 유엔 투표 데이터에 맞는 데이터 마이닝과 비주얼라이제이션 기법을 제안한다. 유엔 회원국의 투표성향은 사물간의 관계를 나타내는 소셜 그래프를 사용하여 시각화할 수 있다. 소셜 그래프는 노드(정점)와 링크(선분)로 구성된다. 노드는 회원국을 의미하며, 두 회원국의 투표성향이 유사하면, 두 회원국 간에 링크가 존재한다. 소셜 그래프를 구현하기 위해서는 회원국 간의 투표성향 유사도(similarity)를 측정하고, 유사도에 따라 그래프를 생성한다. 그리고 그래프 레이아웃 알고리즘을 사용하여, 웹 화면에 보여지게 된다. 그러나 기존의 널리 사용되는 그래프 레이아웃 알고리즘(예: 프루터만-레이골드(Fruchterman-Reingold))을 적용할 경우에 그래프의 높은 복잡도(complexity)로 인해 각 노드의 투표성향을 쉽게 파악할 수 없다(그림 5 참조). 이러한 문제를 해결하기 위해, 본 논문에서는 친구 매칭(Friend-Matching) 알고리즘을 제안하여, 기존 그래프 비주얼라이제이션의 문제점을 개선하였고, 누구나 손쉽게 그래프의 의미를 파악할 수 있도록 하였다(그림 6 참조). 더욱이 기존의 그래프에서 사용하는 링크 정보는 상대적으로 성향이 가까운 노드와 관계만을 사용한다. 이 경우 상대적으로 성향이 다른, 또는 적대적인 국가에 관한 유사도 정보는 활용되지 않는 문제가 존재한다. 이를 해결하기 위해 기존 프루터만-레이골드 알고리즘을 수정하여 친구-라이벌 매칭(Friend-Rival Matching) 알고리즘을 제안하였고, 라이벌 링크를 통해 기존 그래프보다 더 많은 통찰력을 보여줄 수 있었다(그림 8 참조). 또한 기존 그래프는 전반적인 국제관계를 이해하는데 큰 도움을 준다. 그러나 대부분의 경우 하나의 국가가 정해진 상태에서 나머지 국가와의 관계를 보려는 경우에는 적절하지 않다. 이를 개선하기 위해 버블힙(Bubble Heap) 개념을 제안하고 이를 시각화하는 알고리즘을 구현한다. 마지막으로 유엔 투표 데이터를 시각화하여 웹에서 보여주는 프로토타입(http:

//datalab.kunsan.ac.kr/politiz/un/)을 구현하였다. 이를 통해 누구나 웹 페이지를 방문하여 유엔 회원국 간의 투표성향을 파악하고 사용자 인터랙션을 통해 필요한 정보를 얻을 수 있다. 또한 제안된 프로토타입 시스템은 특정 주제(예: 중동 문제 관련 법안들)에 대해 회원국 간의 투표성향을 시각화 하여 회원국 간의 관계를 알아보고, 특정 국가를 쉽게 찾을 수 있는 검색과 하이라이트 기능, 시계열(time-series) 비주얼라이제이션, 노드 크기 조정과 같은 환경 설정 기능을 제공한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 요약, 정리하고, 3장에서는 기존 그래프 비주얼라이제이션을 개선한 친구 매칭, 친구-라이벌 매칭, 버블맵 알고리즘들을 제안하고 그 효용성에 대해 논의한다. 4장에서는 웹에서 동작하는 프로토타입 시스템 구현에 대해서 자세히 기술하고, 5장에서 본 연구에 대한 결론을 내린다.

2. 관련연구

투표 데이터를 기반으로 한 상원의원의 투표성향을 그래프로 표현하고자 하는 연구는 정치학 분야에서 오랜 기간 연구대상이었다[3,6,9]. 정치학에서의 접근 방식은 의안이 포함하고 있는 내용을 모델에 포함시키고 베이시안 이론과 통계학 이론을 접목하여 의원의 성향을 2차원 화면에 배치하는 방식을 주로 사용한다. 이를 위해서는 각개 의안에 대한 키워드 정보가 포함되어야 한다. 의안에 대한 추가적인 정보를 사용하지 않고 투표의 유사도만을 사용한 접근법은 오데완(Odehwan)에 의해 처음 시도되었다[11]. 오데완은 미 상원의원의 투표기록을 바탕으로 1991년부터 2009년까지 9개회기(session)의 상원의원의 투표 성향 비주얼라이제이션을 구현하였다. 9개회기를 분석하였으나 매 회기를 별도의 그래프로 구현하였기에 연속성에 대한 비주얼라이제이션은 포함되지 않았다. 오데완은 투표 데이터를 기반으로 의원 간의 유사도 매트릭스를 계산한 다음 유사도 값이 정해진 컷오프(cut-off) 값보다 큰 경우에만 노드간 링크를 구성하는 방식을 사용하였다. 단순히 유사도 값의 크기를 기준으로 한 링크 선정 알고리즘은 이해하기는 쉬우나 얻어진 결과는 의원 간의 차이를 설명하기에는 부족한 점이 많다. 뷰귀테(Beaugutte)는 유럽연합에 소속한 국가의 관계를 소셜 그래프를 이용하여 비주얼라이제이션 한 예를 발표하였으며, 오데완의 방법과 유사하다[1]. 위의 두 사례는 투표를 한 개인이나 국가를 노드로, 개인이나 국가의 투표 유사도를 링크로 표현한 그래프로 보여주었다는 점에서 본 연구와 유사한 점이 있다. 그러나 오데완의 미 상원이나 뷰귀테의 유럽국가의 관계 표현에 있어 얻어진 그래프는 일반인의 상식과는 다른 결과를 보여주고 있어 링크의 선정 알고리즘이 만족스럽지 못함을 보여준다. 본 연구에서는 노드 수가 제한된 경우 링크의 수를 특정 값으로 제한하여 상식적인 멘탈 모델과 근접한 비주얼라이제이션을 얻는 알고리즘을 제시한다.

또한 [18]은 그래프로부터 커뮤니티를 추출할 수 있는 새로운 링크 클러스터 알고리즘을 제안하였고, 제18대 국회의원들의 투표 데이터를 사용하여 소셜 그래프를 생성하고, 커뮤니티 스트럭처(community structure)를 추출하여 국회의원 투표성향 네트워크에 대한 분석과 비주얼라이제이션을 시도하였다. [17]은 새로운 뉴스 비주얼라이제이션 방법을 제안하였다. 기존의 뉴스로부터 정보를 추출하고, 동일 뉴스기사에 등장한 정보를 서로 관련 있는 것으로 가정하여 그래프를 생성하였다. 즉, 정보를 노드로 구성하고, 정보원 간에 관련이 있으면 링크를 생성하여 그래프를 구성하였다. 1990년부터 국내의 주요 언론사를 포함하는 총 66개 매체로부터 생산된 약 29백만 건의 뉴스기사들로부터 정보원 연결 그래프를 만들고, 뉴스 시각화를 수행하였다. 기간을 설정하고, 키워드를 검색하면, 그 키워드와 관련된 정보원을 찾아서, 그래프를 시계열로 보여준다. 또한 [16]은 국내 주요 언론사의 논조 차이점을 파악하기 위해 소셜 그래프 생성, 분석, 비주얼라이제이션 기법을 제안하였다. 예를 들면, 동아일보, 경향일보 등의 언론사에서 특정 사건과 관련 있는 키워드의 빈도수를 계산하거나 코사인 유사도를 사용하여 키워드 간의 상관관계를 의미하는 그래프를 만들고, 군집화(clustering)를 수행하였다. 최근에 한스 로슬링은 갭 마인더(GapMinder)라는 통계분석 서비스를 일반에 공개하였다[5]. 유엔의 데이터를 바탕으로 3개의 상이한 데이터 세트를 시간에 흐름에 따라 시각화 하였다. 예를 들면, x축에 1인당 GNP, y축에 기대수명을 적용하고, 노드의 크기는 인구에 비례하도록 크기를 조정하였다. 그리고 시간 흐름에 따라 그래프의 변화를 직관적으로 보여주는 시각화 방식을 채택하였다. 최근에는 그래프의 사이즈가 커지면서, 이를 시각화하는 문제에 관한 연구가 활발히 진행되고 있다. [10]에서는 노드의 수가 10억 개 이상이나 되는 대용량 그래프를 시각화하는 새로운 방법을 제안하였다. 특히, 10억 개 노드간의 링크 정보를 화면에 렌더링 하고 특이점(outlier)를 자동으로 발견하여 보여주는 알고리즘을 개발하였다.

기존 소셜 그래프 방식을 UN 투표 법안 데이터에 그대로 적용할 경우에 회원국 간의 전반적인 투표성향이나 클러스터 개수에 대한 분석이 가능하다. 그러나 많은 노드를 비주얼라이제이션하거나 노드간의 링크들이 많아지면, 많은 노드들이 서로 겹치면서 개별 노드에 대한 투표성향을 파악하기 힘들게 된다. 본 논문에서는 친구 매칭 알고리즘을 제안하여 이런 문제를 해결하였다. 또한 기존 소셜 그래프에서는 노드간의 최대적인 투표성향은 알 수 없기 때문에, 친구-라이벌 매칭 알고리즘을 제안하여, 유사한 투표성향을 보이는 노드간에는 청색 링크를 연결하고, 최대적인 투표성향을 보이는 노드간에는 적색 링크를 연결하여 보다 균형 잡힌 시각에서 노드간의 투표성향을 파악하는 것이 가능하다. 마지막으로 버블맵 그래프를 제안하여, 사용자가 관심 있는 한 국가(예를 들면, 한국)에 대해서 나머지 회원국들과의 투표성향 유사 정도를 쉽게 알 수 있다. 기존 소

설 그래프는 모든 노드에 대한 전체적인 투표성향을 파악하기는 용이하나, 특정 노드에 대한 나머지 노드의 투표성향을 파악하는 것은 쉽지 않다.

3. 제안방안

1) UN 회원국 간의 유사도 측정

에릭 보텐(Erik Voeten)의 데이터베이스[15]에서 UN 총회의 투표 데이터를 데이터베이스 형태로 일반에 공개하고 있다. UN 총회에서는 모든 의안을 투표에 상정하는 것이 아니며 상당수의 의안은 조율을 통해 만장일치로 처리되며 회원국 간에 의견이 갈리는 경우 투표로 결정한다. 2012년 말까지 투표로 처리한 법안 수는 5,211건으로 연간 77건 수준이다. 전체 기간을 1년 단위로 분석할 경우 1946년부터 2012년까지의 67개 구간으로 지나치게 많다. 따라서 해당 기간을 5년 단위 기간(term)으로 나누어 분석하였다. 의안은 일반적으로 투표에 찬성(in favour)과 반대(against) 의견을 표시하나 기권(abstain)과 불참(absent)으로 투표에 참석하지 못할 수 있으며, 회원국 A와 B 간의 유사도는 두 회원국이 같은 의견을 표시한 투표 횟수를 두 회원국이 동시에 투표에 참여한 횟수로 나눈 값으로 정의한다.

$$A \text{와 } B \text{의 유사도 (sim(A,B))} = \frac{A \text{와 } B \text{가 찬성 또는 반대 의견을 함께 표시한 횟수}}{A \text{와 } B \text{가 법안에 대한 찬성 또는 반대로 의견을 표시한 횟수}} \quad (\text{수식 1})$$

위와 같이 두 회원국 간의 투표성향 유사도를 계산한다.

2) 그래프 레이아웃(layout) 알고리즘

그래프를 화면에 렌더링(rendering)하기 위해서 니토(neato)라는 알고리즘을 사용한다. 니토 알고리즘은 노드를 양전하로 충전된 입자로 보고 노드간의 연결 에지를 스프링으로 모델링하여 같은 전하로 충전된 노드간의 척력(斥力)과 스프링으로 연결되어 끌어 다니는 인력(引力)의 평형점을 찾도록 하는 것이다. 노드의 배치에 있어 니토 외에도 프루터만-레인골드 알고리즘도 함께 사용한다[4]. 프루터만-레인골드 알고리즘은 니토의 스프링 모델과 유사하나 평형점을 찾는 것을 목표로 하는 대신 링크로 연결된 노드간에는 인력이 작용하며 링크가 없는 모든 노드간에는 척력이 작용하는 모델을 사용한다. 처음에는 노드의 위치를 랜덤으로 정한 다음 노드간의 거리에 따라 결정되는 외력을 가해 새로운 위치에 노드를 배치한다. 이러한 사이클을 반복하는 동시에 한 사이클에서 이동할 수 있는 거리를 점진적으로 축소하여 노드의 안정된 위치를 찾는 방식이다. 이때 노드간의 인력과 척력은 그래프 면적과 상관되는 상수 k 를 사용하여 정의한다. k 는 노드 주위에 빈 공간(empty area)의 반경(radius)를 나타내고 d 는 두 노드간의 거리(distance)를 나타낸다. 상수 C 의 값은 실험을 통해 적당한 값을 사용하며 인력(f_a)과 척력(f_r)은 다음과 같이 계산한다.

$$k = C \sqrt{\frac{\text{그래프 면적}}{\text{노드들의 개수}}}$$

$$f_a = \frac{d^2}{k}$$

$$f_r = -\frac{k^2}{d}$$

이 두 알고리즘은 실제 레이아웃 결과에서 큰 차이를 보이지 않는다[4]. 본 연구에서는 프루터만-레인골드 알고리즘을 기본으로 사용한다.

그림 5는 기존의 오데완과 뷰귀테 방식에 의해 최근 2년(2011-12) 간의 투표성향을 유사도 컷오프 0.65 기준으로 그린 그래프이다. 이 그래프를 보면 미국, 캐나다와 일부 태평양 국가의 클러스터, 유럽국가 클러스터, 나머지 대부분의 국가 클러스터 등 3개의 클러스터가 만들어짐을 볼 수 있다. 클러스터를 확인할 수 있다는 것은 바람직하지만 클러스터 내부의 회원국 간의 관계를 보고자 하는 경우에는 도움이 되지 않는 것을 알 수 있다.

3) 친구 매칭 알고리즘

그림 5에서 보는 바와 같이 기존 방식대로 그래프를 화면에 디스플레이 하면 회원국들은 세 개의 클러스터로 나누어지고 몇 개의 아웃라이어를 발견할 수 있으나 나머지 대다수의 회원국의 투표성향을 파악하는 데는 큰 도움이 되지 않는다. 클러스터 내의 회원국들은 많은 링크들로 서로 연결되어 있기 때문에, 클러스터 내의 특정 노드의 투표성향을 파악하기는 어렵다. 따라서 링크 수를 줄여 클러스터 내부를 볼 수 있도록 하는 것이 하나의 방법이 될 수 있다. 이러한 아이디어를 채택한 새로운 알고리즘은 다음과 같이 제안한다.

노드간의 링크는 쌍방향성이기에 노드간에 링크를 구성하려면 해당 노드에서 보는 유사도의 값이 양쪽 모두에게서 상대적으로 높을 값이어야 한다. 노드 당 n 개의 링크를 구성한다고 하자. 모든 노드에 대해 각 노드에서 보는 유사도 값이 n 번째 이내의 링크 모두를 1차적으로 선택한다. 이렇게 선택된 링크가 쌍방에서 함께 선택된 링크일 경우 이를 최종적으로 선정한다. 예를 들어, A, B, C, D의 유엔 회원국 중에서 A와 투표성향이 유사한지를 (수식 1)을 통해 다음과 같이 구했다고 가정하자. $\text{sim}(A,B) = 0.9$, $\text{sim}(A,C) = 0.7$, $\text{sim}(A,D) = 0.6$. 이 중에서 A와 유사한 투표성향을 보인 회원국 중에서 두 개($n=2$)를 선택하면, 유사도 값이 높은 순서대로 B와 C가 될 것이고 이들을 A의 친구(friend)라고 한다. 마찬가지로 B와 유사한 투표성향을 다음과 같이 구했다고 가정하자. $\text{sim}(B,A) = 0.9$, $\text{sim}(B,C) = 0.5$, $\text{sim}(B,D) = 0.8$. 이 경우에 B의 친구는 A와 D가 된다. B는 A의 친구이고, A는 또한 B의 친구이기 때문에, A와 B를 링크로 연결한다. 기존 방식에서는 유사도가 특정 임계값(threshold value)보다 높으면, 두 노드 사이를 링크로 연결하지만, 친구

매칭 알고리즘에서는 두 노드가 서로 친구일 경우에만 선분을 연결한다. 만일 노드 당 원하는 수의 링크가 만들어지지 않을 경우 선택된 링크를 제거하고 같은 작업을 반복한다. 예를 들면 특정 노드에게 2개의 링크가 선택되었다면 다음 사이클에는 n-2개의 링크만을 선택하는 방식으로 링크 개수를 늘려나가며 모든 노드가 n개의 링크를 구성하거나 남아있는 노드의 최댓값이 정해진 컷오프 값 이하로 될 때까지 이러한 사이클을 반복한다.

이와 같은 알고리즘은 반에서 친구를 정하는 방식과 비슷하다. 한 반의 학생을 50명이라고 할 때 모두를 친구로 하기는 어렵고 대략 10명 이내의 수를 친구로 한다. 이때 친구를 선택하는 방식과 유사하기에 알고리즘을 친구 매칭이라 정의한다. 친구 매칭 알고리즘을 사용하고자 할 경우 적절한 친구의 수가 몇 명이 적당한지를 판단하기 위해 링크의 수를 변화할 경우 그래프의 모양에 차이가 있는지를 확인하였다. 실험 결과에 따르면 친구 수 (n)이 10 이상일 경우 전체적인 모양의 변화는 많지 않으며 n이 증가할 경우 클러스터의 중심부에 밀도가 높은 클러스터가 만들어지는 것을 볼 수 있다. 이러한 현상을 기반으로 n=10을 일반적인 분석에 사용하도록 한다.

같은 유사도 데이터에 친구 매칭 알고리즘을 적용할 경우 그림 6의 그래프를 얻을 수 있다. 그림 5와 비교하면 3개의 클러스터에 관한 내용을 포함해서 여러 국가 간의 관계가 상대적으로 이해하기 쉽게 배치되어 있음을 확인할 수 있다.

4) 친구-라이벌 매칭 알고리즘

친구 매칭 알고리즘의 기본원리는 노드의 링크 개수를 적당한 수준으로 유지할 경우 특정 컷오프 값 이상의 모든 링크 정보를 활용하는 것보다 이해하기 쉬운 그래프를 얻을 수 있다는 것이다. 이 때 사용하는 링크 정보는 상대적으로 성향이 가까운 노드와의 관계만을 사용한다. 이 경우 상대적으로 성향이 다른, 또는 적대적인 국가에 관한 유사도 정보는 활용되지 않는다. 예를 들어 미국과 이스라엘의 유사도는 미국과 다른 국가와의 유사도 값보다 상대적으로 크며 결과적으로 미국과 이스라엘은 가까운 위치에 배치된다. 그러나 이스라엘과 아랍 국가와의 관계처럼 총회에서 적대적인 투표활동을 한 정보는 활용되지 않았다. 친구 매칭 알고리즘으로 얻어지는 그래프에서 적대적인 국가와는 상대적으로 먼 곳에 배치되지만 적대적이라는 정보를 활용할 보다 개선된 그래프를 얻을 수 있을 것이다. 친구 매칭 알고리즘을 확대하여 쌍방이 낮은 유사도 값을 갖는 국가를 링크로 선정하는 알고리즘을 친구-라이벌 매칭 알고리즘이라 정의한다. 친구 링크와 구별하기 위해 라이벌 관계로 얻어지는 링크는 라이벌(rival) 링크로 정의한다.

앞서 노드 배치에 적용한 프루터만-레인골드 알고리즘은 링크로 연결된 노드와의 인력과 링크로 연결되지 않은 노드와의

척력 두 가지 힘만을 사용하여 노드를 배치한다. 친구-라이벌 매칭 알고리즘을 적용하기 위해서는 프루터만-레인골드 알고리즘을 일부 수정하여야 한다. 친구 링크의 경우 기존 프루터만-레인골드 알고리즘의 인력 공식을 사용하나 라이벌의 링크의 경우에는 인력 공식을 수정하여 노드가 보다 먼 곳에 위치하도록 한다. 이를 반영하기 위해 인력을 거리에 반비례하기보다 거리의 제곱근에 반비례하는 방식을 사용한다. 링크의 존재여부와 상관없이 모든 노드간의 척력은 프루터만-레인골드 알고리즘의 공식을 그대로 사용한다. 이를 정리하면 다음과 같다.

- 친구 링크로 연결된 노드간의 인력: $f_f = \frac{d^2}{k}$
- 라이벌 링크로 연결된 노드간의 인력: $f_r = \frac{k^2}{\sqrt{d}}$
- 모든 노드간의 척력: $f_o = -\frac{k^2}{d}$

그림 7은 특정 기간(1976-80년)의 유엔 투표성향을 친구매칭과 친구-라이벌 매칭 알고리즘을 적용한 결과를 보여준다. 우측 상단에 위치한 동구권 국가와 우측 하단의 서방측 국가의 클러스터의 상대적 위치는 비슷하나 친구-라이벌 매칭을 적용할 경우 동구권 국가가 서방 국가의 위치에 상대적으로 먼 위치에 배치된 것을 볼 수 있다. 청색의 친구 링크와 적색의 라이벌 링크를 볼 수 있으며 이러한 링크 정보를 통해 국가 간의 이해관계를 파악하는데 도움이 됨을 볼 수 있다. 1976-80년에는 동서 진영의 대립과 함께 제 3세계 국가의 등장으로 인해 다극화된 국제관계에서는 친구 매칭보다 친구-라이벌 매칭이 라이벌 링크를 통해 더 많은 통찰력을 보여줄 수 있다.

5) 버블힘 알고리즘

앞에서 기술한 유사도와 매칭 알고리즘을 사용한 비주얼라이제이션은 유엔 총회 투표에서의 유사성을 비교한 것으로 국가로 표시되는 노드의 위치가 인접해 있는가를 기준으로 판단할 수 있으며 전반적인 국제관계를 이해하는데 도움이 된다. 그러나 대부분의 경우 하나의 국가가 정해진 상태에서 나머지 국가와의 관계를 보려는 경우가 많다. 이러한 정보는 2차원 정보가 아니기에 유사도 매트릭스의 특정 행이나 열 내용에서 직접 확인할 수 있으나 200여개 국가의 유사도를 한눈에 파악하기는 쉽지 않다. 더구나 국가 간 관계를 파악함에 있어 강대국이나 선진국과의 관계는 개발도상국이나 인구가 적은 소국과의 관계에 비해 중요하기 때문에 이러한 정보를 쉽게 볼 수 있는 비주얼라이제이션 기법이 필요하다.

그림 9(a)은 x축만을 사용하여 특정 국가의 유사도를 표시한 그래프이다. 이 때 다른 국가는 같은 크기의 원으로 표시하지 않고 GDP나 인구에 비례한 원으로 표시하였다. x축 하나만을 사용하고 있어 원으로 그려지는 노드가 중복되어 그려지며 결과적으로 그래프가 전달하고자 하는 정보를 이해하기에는

어려운 점이 있다. 노드의 수가 증가한다면 중복현상은 심화되어 그래프의 의미를 파악하기 어렵다. 그래프에서 중복되는 노드를 중복되지 않도록 y축 방향으로 이동한다면 중복을 방지할 수 있다. 다만 많은 노드가 중복될 경우 x값을 그대로 유지하고 y값만을 변경할 경우 지정된 공간 안에 그래프를 그리는 것이 불가능한 경우가 있을 수 있다. 이때 x좌표의 값을 약간이라도 변경할 수 있도록 제약을 완화하면 전체적인 정보를 쉽게 파악할 수 있으며 이러한 방식을 버블힙이라 정의한다.

버블힙 비주얼라이제이션에서 노드 배치 알고리즘은 노드의 위치를 수정하여 안정적인 위치에 놓일 때까지 다음과 같은 반복 과정을 거친다.

- 0) 노드 i의 초기 위치, $x_i = \text{유사도}$, $y_i = \text{random}$ 으로 정한다. (random=0~1의 랜덤 값)
- 1) 노드의 다음 좌표는 다음과 같은 외력을 합해서 결정된다.
 - a. 좌표($x_i, 0$)에서의 인력
 - b. 타 노드와 충돌이 있을 경우 충돌에 의한 반작용
- 2) 노드의 최대 이동거리는 반복할 경우 작아지도록 쿨링(cooling) 함수를 적용하여 시간이 지나면 더 이상 변하지 않도록 한다.

그림 9(b)는 버블힙이 진행되는 중간 결과로 버블로 표시되는 노드 사이에 공간 여유가 있으나 그림 9(c)의 경우 안정적인 위치에 배치된 것을 볼 수 있다. 버블힙 알고리즘은 모든 노드의 유사도 값과 실제 배치된 x좌표 값이 정확히 일치하지는 않으며 이를 최소화하기 위해서는 여러 가지 변형된 충돌 모형을 사용할 수 있다. 드물게 유사도 값이 0이나 1과 같은 실수 값을 갖는 경우들이 나타난다. 이 경우 여러 노드가 좌우로 이동할 여지가 없어 상하 직선상에 놓이는 현상이 발생할 수 있어 노드의 초기 위치를 정하는 1)단계를 다음과 같이 일부 수정해야 한다.

- 1a) 노드 i의 초기 위치, $x_i = \text{유사도} + C \times \text{random}$, $y_i = \text{random}$ 으로 정한다. C는 0.001과 같은 작은 수로 x좌표의 값에 같은 값이 생성되지 않도록 노이즈 삽입 기능을 말한다.

우리나라의 기준으로 본 가장 최근의 외국과의 관계는 그림 9(c)와 같다. 그림에서 보면 큰 원으로 표시된 강대국의 경우 일본, 독일과 투표성향이 유사하며 중국, 인도와는 상대적으로 다른 것으로 보인다.

6) 제안 알고리즘의 특징 요약

각 제안방안을 통해 얻어진 그래프에서 노드간 거리, 노드간 링크 의미, 노드 크기, 노드 위치에 대한 의미를 표 1에서 정리한다. 또한 기존 그래프에서 나타나지 않는 현상을 설명한다.

표 1. 제안 알고리즘의 특징 요약

특징	친구 매칭	친구-라이벌 매칭	버블힙 그래프
노드간 거리	근거리 (투표성향 유사)	근거리 (투표성향 유사)	의미 없음
노드간 링크 의미	투표성향 유사	투표성향 유사(정색), 적대적인 투표성향 (적색)	링크 없음
노드 크기	인구 또는 GDP	인구 또는 GDP	인구 또는 GDP
노드 위치	의미 없음	의미 없음	x축: 유사도, y축: 의미 없음
기존 그래프와 다른 점	클러스터 내부에 중첩된 노드의 투표성향 파악 가능	적대적인 투표성향 파악 가능	특정 노드에 대한 다른 노드의 투표성향 파악용이

7) 제안 알고리즘의 유효성 검증

본 논문에서 제안한 그래프 비주얼라이제이션 알고리즘의 유효성을 검증하기 위해 사용자 테스트를 수행하였다. 총 39명을 대상으로 성별, 나이 등의 기본 정보와 데이터 비주얼라이제이션에 대해 알고 있는지를 물어보았다. 또한 공공데이터 개방 및 시각화에 대해 잘 알고 있는지, 국제 외교에 관심이 있는지를 조사하였다. 그리고 냉전, 독일 통일 등 기본적인 외교 상식을 테스트한 다음, 기존 방안과 제안 방안을 통해 만들어진 그래프들을 보여주고, 어느 그래프가 좀 더 효과적인지를 물어보았다.

조사 대상자 중 남성은 27명, 여성은 12명이었고 나이는 만 18세~만 28세였다. 전공은 통계전공 17명, 컴퓨터전공 22명으로 경기대, 고려대, 군산대, 아주대, 한국과학기술원에 재학 중인 대학생 및 대학원생이었다. 표 2는 본 논문과 관련된 주제에 대한 설문 응답자들의 기초적인 지식수준에 대한 통계 결과를 보여준다.

표 2. 설문 응답자들의 배경 지식에 대한 통계 결과

문항	예	아니오
데이터 비주얼라이제이션에 대해 들어본 적이 있는가?	28명 (72%)	11명 (28%)
공공데이터가 무엇인지 알고 있는가?	15명 (38%)	24명 (62%)
평소 외교와 유엔 회원국의 법안 투표에 관심이 있는가?	1명 (3%)	38명 (97%)
냉전(cold war)에 대해 알고 있는가?	29명 (74%)	10명 (26%)

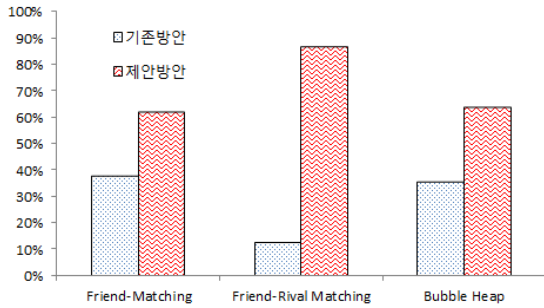


그림 1. 제안방안의 유효성 테스트 결과

그림 1에서 보는 것처럼, 응답자의 60% 이상이 기존방안보다 제안방안에 의해 얻어진 그래프를 통해 유엔 회원국 간의 투표성향을 잘 알 수 있다고 대답했다. 친구 매칭 알고리즘을 테스트하기 위해, 그림 5(기존방안)과 그림 6(제안방안)을 제시하고 한국(KOR)을 찾고, 한국과 다른 유엔 회원국 간의 투표성향을 잘 보여주는 그림을 찾으라고 하였다. 62%의 응답자는 제안방안을 선택했는데, 그 이유는 그림 5에서는 노드들이 중첩되어 있어 한국을 찾을 수 없기 때문에 한국과 다른 회원국 간의 투표성향을 파악할 수 없다고 응답하였다. 그러나 38%는 기존방안을 통해 얻어진 그래프가 더 효과적이라고 응답했다. 그 이유는 한국을 찾을 수는 없지만 클러스터 간의 관계를 명확히 알 수 있는 것은 그림 5라고 대답하였다. 친구-라이벌 매칭 알고리즘 테스트에서는 87%의 응답자가 제안방안에서 얻어진 그림 8(제안방안)이 좀 더 적대적인 관계를 나타낸다고 응답했다. 그림 7(기존방안)과 8(제안방안)은 1976-80년에 UN 법안 투표 결과를 시각화 한 그래프이고, 이때는 냉전 시기(1) 미국/서유럽을 중심으로 하는 자유 진영, (2) 소련/동유럽의 공산권, 그리고 (3) 인도/동남아시아/아프리카 등의 제3세계로 나뉘어 대립하던 시기였다. 그림 8은 회원국 간 유사한 투표성향은 청색 링크로, 적대적인 투표성향은 적색 링크로 표시되어, 응답자들이 쉽게 적대적인 관계를 유추할 수 있던 것으로 판단된다. 마지막으로 버블힙 그래프 테스트에서는 64%의 응답자가 기존의 소셜 그래프 방식(그림 5)보다 유용하다고 응답했다. 그 이유는 소셜 그래프가 모든 회원국 간의 전체적인 투표성향을 파악하기는 용이하나, 특정 국가에 대해 다른 국가 간의 투표성향을 파악하기 힘들기 때문이다.

4. 프로토타입 구현

1) 프로토타입 레이아웃

그래프 비주얼라이제이션을 위해서는 링크 구성을 위한 투표 데이터 외에 노드를 표시하기 위한 회원국 데이터를 확보하여야 한다. 회원국을 표시할 경우 2자 또는 3자의 알파벳을 사용한다. 국제 올림픽 위원회(IOC), 국제축구연맹(FIFA)와 같이 전 세계의 국가를 대상으로 하는 기관은 나름대로 독자적인 국가코드를 사용하고 있다. 본 논문에서 국가코드는 국제표준

기구(ISO)의 ISO 3166-1 alpha-3을 사용한다[8]. 회원국 수는 67년이라는 기간 동안 변화가 있으며 이에 대한 이해가 필요하다. 신규로 가입하는 국가는 당연한 것으로 문제가 되지 않으나 동서독의 통일, 소련연방의 해체, 유고슬라비아 해체와 같은 경우에는 국가의 분할이나 통합의 경우 우리가 생각하는 멘탈 모델에 맞는 합리적인 처리가 필요하다. 독일의 경우에는 동독이 서독으로 흡수통합으로 서독의 데이터를 독일로 함께 사용하였으며 체코슬로바키아의 경우에는 체코와 슬로바키아의 두 국가로 분리 표시되며 체코슬로바키아는 별도의 국가로 정의한다. 소련연방 해체의 경우에는 러시아가 소련을 이어받도록 처리하였으며 독립된 연방은 새로운 국가명을 부여하였다. 유고슬라비아의 경우에는 세르비아가 유고슬라비아를 계승하는 것으로 처리하였다. 또한 유엔 회원국으로 가입은 했으나 실제 투표에 참여한 횟수가 아주 적은 적도 기니(República de Guinea Ecuatorial, GNQ), 키리바티(Kiribati, KIR), 남 수단(South Sudan, SSD)과 같은 국가는 데이터 부족으로 그래프에 표시되지 않았다.

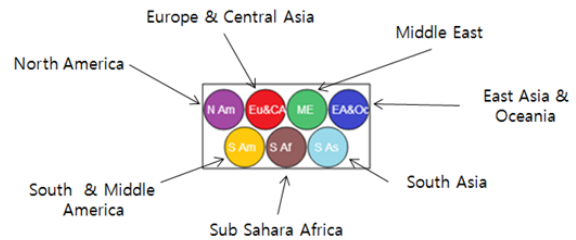


그림 2. UN 회원국의 지역별 컬러 코딩

특정 국가의 이해관계는 국가가 속한 대륙이나 지역과 관계가 있어 지역 표시가 필요하다. 이러한 지역적인 구분을 위해서는 단순 대륙별 구분을 사용하지 않고 월드뱅크에서 사용하는 대륙별 구분을 일부 변형한 그림 2에서 표시한 방식을 사용하였다.

이러한 지역 구분을 기초로 국가별 컬러를 다르게 표시하면 비주얼라이제이션에서 지역 간에 어떤 유사성이 있는지를 쉽게 인지할 수 있다. 국가의 규모를 대략이라도 표시하고자 할 경우 GDP 데이터를 참고로 사용하였다. 이러한 표시방법을 정한 후 앞서 제시한 여러 알고리즘을 이용하여 회원국 간의 유사성을 그래프로 표시하였다. 그리고 유사한 법안들을 그룹으로 묶고 특정 주제(예: 중동 문제 관련 법안들)에 대해 회원국 간의 투표성향을 알 수 있다. 또한 특정 국가를 쉽게 찾을 수 있는 검색과 하이라이트 기능, 시계열 비주얼라이제이션, 노드 크기 조정과 같은 환경 설정 기능을 제공한다.

2) 그래프 렌더링 구현

앞서 설명한 샘플 그래프를 얻기 위해 두 가지 개발 도구를

사용하였다. 유사도로부터 친구 매칭이나 친구-라이벌 매칭 알고리즘을 사용하여 링크를 추출하는 프로그램은 엑셀에 내장되어 있는 Visual Basic Script 프로그램을 이용하였다. 웹 서비스 전의 프로토타입은 HTML5와 HTML5의 그래픽 기능인 캔버스(canvas), HTML5의 프로그램 언어인 JavaScript 프로그램을 사용하였다.

알고리즘의 테스트를 위한 비주얼라이제이션은 시간 개념이 포함되어 있기에 알고리즘 구현에 추가적인 휴리스틱(heuristic)이 필요하였다. 노드를 캔버스에 랜덤 하게 배치하고 노드 좌표를 알고리즘에 따라 새로 계산하고, 계산된 변화값의 최대치를 쿨링 함수를 사용하여 줄여 나가며 수십 번 반복 계산하면 노드 좌표의 안정적인 값을 얻을 수 있다. 보드 배치에 사용된 프루터만-레인폴드 알고리즘의 단점은 초기에 랜덤함수를 사용할 경우 얻어지는 그래프는 매번 다르게 그려진다는 것이다. 전반적인 형태는 거의 같으나 좌우 반전이 있거나 회전과 같은 좌표 변화를 해야 같은 모양을 얻을 수 있다. 특히 매 기간마다 이전의 좌표에서 새로운 좌표로 노드가 이동하는 경우 노드의 배치에 일정한 규칙이 정해지지 않으면 애니메이션을 이해하기 어렵게 된다. 이를 위해서 시간 t의 그래프를 G(t)라 할 때 다음과 같은 기준을 만족하여야 한다.

- 0) G(t + 1)의 그래프의 노드 좌표는 G(t) 그래프의 노드 좌표에서 가급적 적게 변해야 한다.
- 1) G(0)의 그래프는 항상 동일한 형태로 각 노드의 좌표는 프로그램을 반복 시행할 경우 같은 위치에 배치되어야 한다.

1)을 만족하기 위해서는 G(t + 1)의 노드 좌표는 G(t)의 노드 좌표로부터 출발하여야 함을 말한다. 브랜드스(Brandith)[2]는 일련의 그래프를 얻고자 할 경우 이전 노드의 좌표로부터 인력을 적용하면 그래프의 안정성을 개선할 수 있는 앵커링(anchoring) 개념을 제시하였으며 본 연구의 비주얼라이제이션에 같은 개념을 포함하였다. 앵커링이란 노드의 다음 좌표의 위치를 결정하고자 할 경우 현재 노드의 좌표로부터 인력이 작용하도록 프루터만-레인폴드 알고리즘을 수정함으로 가능하다.

두 번째 기준은 반복되는 그래프를 구현하며 발견된 문제로 노드의 초기 위치 선정이 그래프의 배치에 영향을 주기 때문에 고정된 위치에서 출발해야 함을 확인하였다. 이에 대한 휴리스틱 해법은 모든 노드를 캔버스 크기의 동심원 상에 일정한 간격으로 배치하여 해결하였다. 노드 번호 1번을 원상의 어느 위치에 배치하는 가에 따라 초기 그래프의 회전이 결정된다. 그림 10은 서방국의 클러스터를 우측에 배치하기 위해 실험적으로 결정한 노드의 초기 배치이다. 현재의 회원국 193개 국가 중 1950년 이전에 유엔에 가입한 국가만이 표시되어 있다.

초기 회원국이 아닐 경우 초기 위치 선정은 가급적 기준 그

래프의 구조에 영향을 적게 미치는 곳에 배치하여야 한다. 캔버스의 중심 위치가 타당해 보이나 여러 노드가 가까운 위치에 배치될 경우 매우 큰 척력이 생성될 수 있기에 작은 반지름을 갖는 원 상에 배치하여 문제를 해결하였다. 그림 11은 1956-70년에 추가로 가입한 회원국의 초기 위치가 작은 원 형태로 배치된 것을 볼 수 있다. 해당 기간에 가입한 회원국만이 배치되었기에 완전한 원이 아니라 불완전한 원 형태로 보인다.

3) 웹 기반의 프로토타입 구현

그림 3은 알고리즘을 테스트하기 위한 프로토타입 프로그램 실행화면을 보여준다. 웹의 경우 프로토타입에서 사용된 알고리즘을 직접 구동할 수도 있으나 사용자의 컴퓨터의 화면 크기를 최대한 활용하기 위해서는 사용자의 스크린에 다른 모양의 그래프가 그려지게 되며 이는 불필요한 혼란을 야기할 수 있어 스크린 좌표를 읽어 웹 브라우저에서 스크린 크기에 맞추어 렌더링 하는 방법을 사용하였다. 그림 4는 웹 비주얼라이제이션을 캡처한 화면을 보여준다. 본 연구에서 개발된 프로토타입은 <http://datalab.kunsan.ac.kr/politiz/un/> 웹사이트에서 현재 일반에게 공개되고 있으며, 사용자는 웹사이트를 방문하여 프로토타입을 사용할 수 있다.

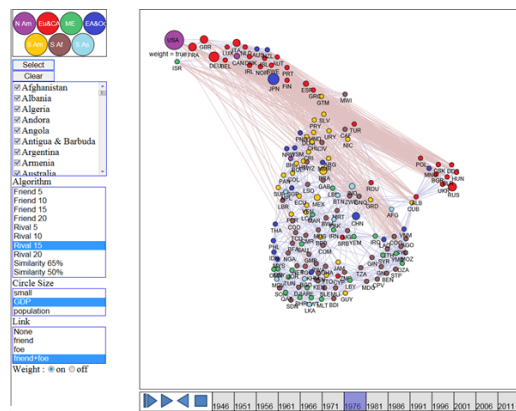


그림 3. 알고리즘 테스트를 위한 프로토타입 화면

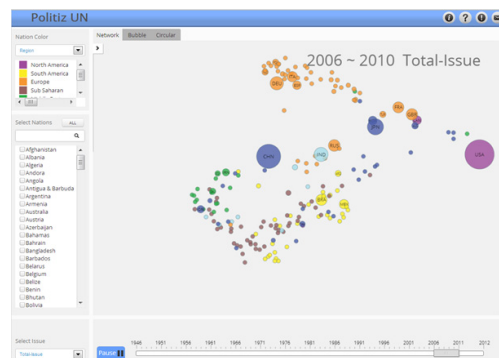


그림 4. 웹 비주얼라이제이션 구현 화면

5. 결론

본 논문에서는 공공데이터 비주얼라이제이션의 중요성을 널리 전파하기 위해 일반인에게 큰 관심이 있고 데이터를 손쉽게 얻을 수 있는 유엔 투표 데이터를 비주얼라이제이션 하여 사용자가 쉽게 이해할 수 있도록 친구 매칭, 친구-라이벌 매칭, 버블형 비주얼라이제이션 기법들을 새로 제안하고 그 효용성을 보였다. 또한 프로토타입 시스템을 개발하여 <http://datalab.kunsn.ac.kr/politiz/un/> 웹사이트에 공개하여 일반 사용자들이 직접 사용할 수 있도록 하였다. 이러한 연구를 통해 유엔 투표 데이터와 같이 사물간의 관계를 표현하기에 용이한 공공데이터에 대한 적절한 비주얼라이제이션 기법을 개발하였으며, 일반 사용자들이 쉽게 이해하고 활용할 수 있는 계기를 만들었다.

참고문헌

- [1] Beauguitte, L. Looking for European Union in the word-system: A multi-graph approach. European Regional Science Association Conference. Available at <http://www.sre.wu.ac.at/ersa/ersaconfs/ersa10/ERSA2010finalpaper698.pdf>. 2010.
- [2] Brandes, U. and Wagner, D. A Bayesian paradigm for dynamic graph layout. Graph Drawing. Springer Berlin Heidelberg. 1997
- [3] De Leeuw, J. Principal component analysis of binary data by iterated singular value decomposition. Computational Statistics & Data Analysis 50(1): 21-39. 2006.
- [4] Fruchterman, T. and Reingold, E. Graph drawing by force-directed placement. Software: Practice and Experience 21(11): 1129-1164. 1991.
- [5] GapMinder. Available at <http://www.gapminder.org>. 2014.
- [6] GovTrack.us: Tracking the U.S. Congress. Available at <http://www.govtrack.us>
- [7] Herman, I. Graph visualization and navigation in information visualization: A survey. IEEE Transactions on Visualization and Computer Graphics 6(1):24-43. 2000.
- [8] ISO 3166-1 alpha-3. Available at http://en.wikipedia.org/wiki/ISO_3166-1_alpha-3.
- [9] Jakulin, A. and Buntine, W. Analyzing the US Senate in 2003: Similarities, networks, clusters and blocs. Available at http://kt.ijs.si/aleks/Politics/us_senate.pdf. 2004.
- [10] Kang, U., Lee, J., Koutra, D. and Faloutsos, C. Net-Ray: Visualizing and mining billion-scale graphs. Pacific-Asia Conference on Knowledge

Discovery and Data Mining (PAKDD). Taiwan, 2014.

- [11] Odewahn, A. Visualizing the U.S. Senate social graph (1991-2009). Beautiful Visualization. O' Reilly. pp.123-142. 2010.
- [12] OECD. Better life initiative, your life index. Available at <http://www.oecdbetterlifeindex.org>.
- [13] Open Knowledge Foundation. Europe's energy. Available at <http://energy.publicdata.eu/ee/vis.html>.
- [14] Open Knowledge Foundation. Where does my money go?. Available at <http://wheredoesmymoneygo.org/dashboard>.
- [15] Voeten, E. and Merdzanovic, A. United nations general assembly voting data. Available online on 2010.
- [16] 감미아, 송민. 텍스트 마이닝을 활용한 신문에 따른 내용 및 논조 차이점 분석. 지능정보연구. 제18권 3호. pp. 53-77. 2012.
- [17] 박대민, 김기남, 강남용, 서봉원, 하효지, 온병원, 저널리즘 가치에 기초한 알고리즘을 이용한 뉴스 시각화. 한국HCI학회 논문지. 제9권 2호. 한국HCI학회. pp. 5-12. 2014.
- [18] 온병원, 이인규, 이만재. Link structure based community detection 알고리즘의 제안과 소셜 네트워크 분석 및 비주얼라이제이션을 위한 사례 연구. HCI2012 학술대회 논문집. 하이원 리포트. pp. 294-297. 2012.



이 만 재

1986년 텍사스대(오스틴) 컴퓨터공학과 박사. 2005년 한국전자통신연구원 디지털콘텐츠연구소 단장. 2011년 차세대융합기술연구원 특임연구위원. 관심분야는 콘텐츠, HCI, 데이터 비주얼라이제이션



온 병 원

2007년 펜실베이니아주립대 컴퓨터공학과 박사. 2008년 브리티시컬럼비아대 포스닥 연구원. 2011년 차세대융합기술연구원 선임연구위원. 2014년 군산대 통계컴퓨터학과 조교수. 관심분야는 데이터 마이닝, 데이터베이스, 정보검색, 빅데이터

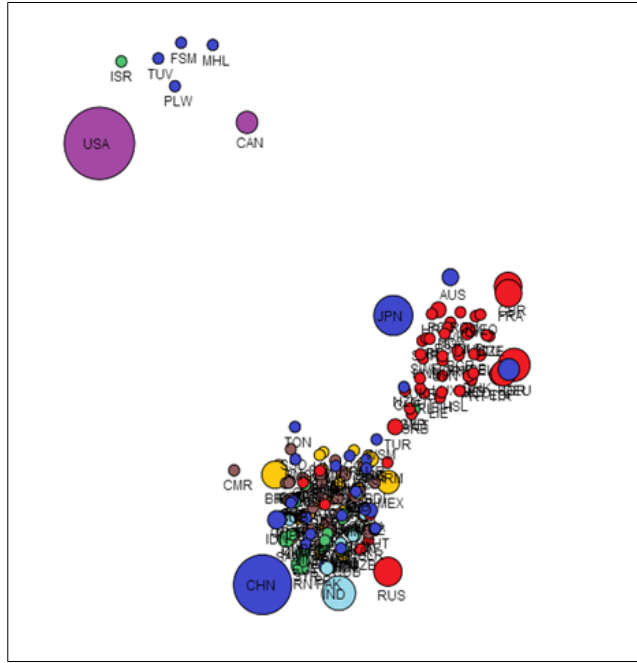


그림 5. 2011-12 UN 회원국 투표성향 (유사도 0.65 기준)

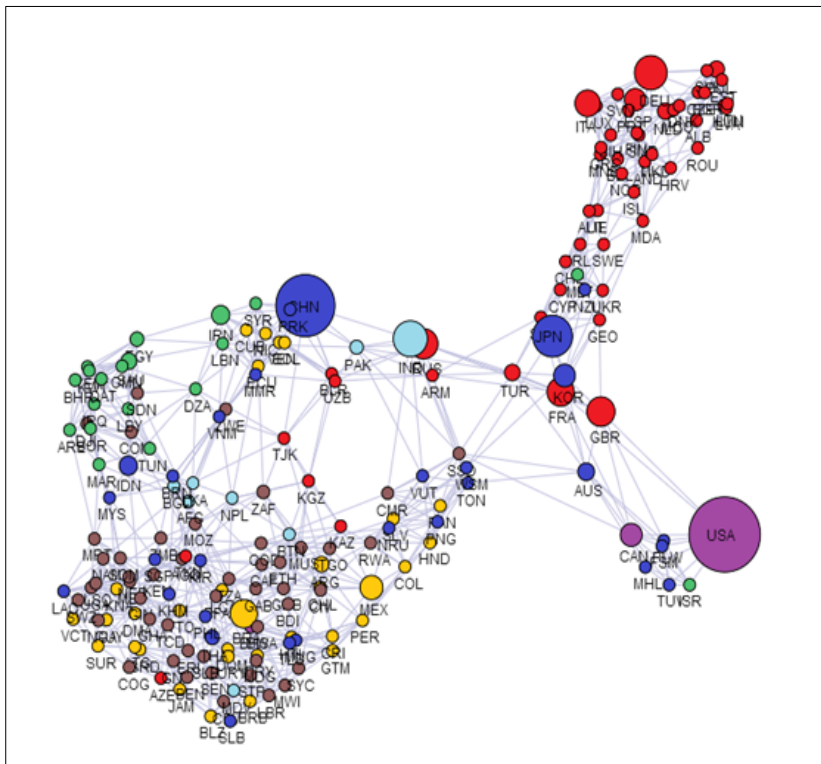


그림 6. 2011-12 UN 회원국 투표성향 (친구 매칭 n=10)

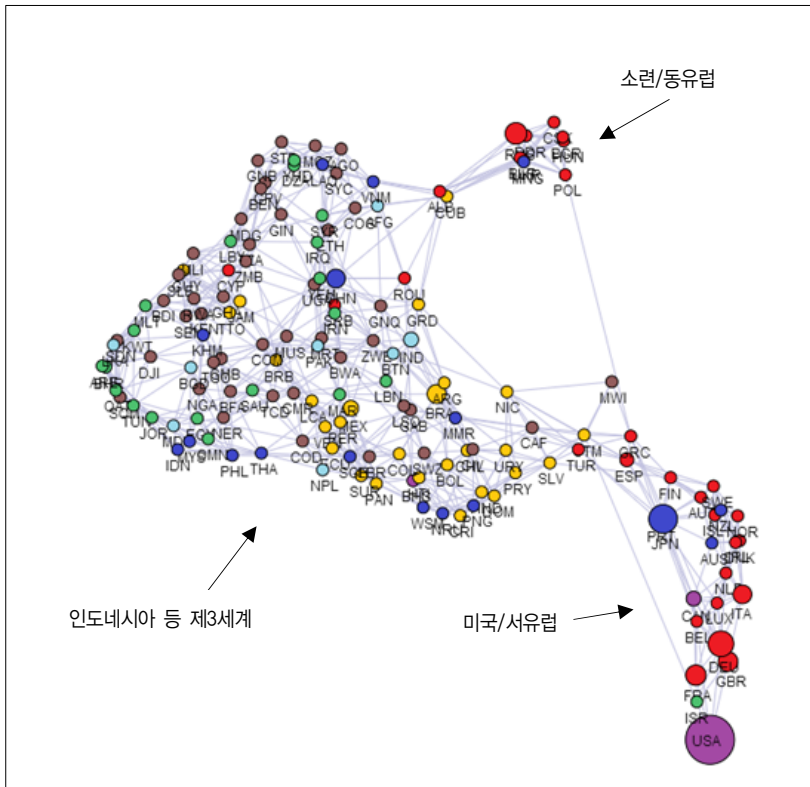


그림 7. 1976-80 UN 회원국 투표성향 (친구 매칭 n=10)

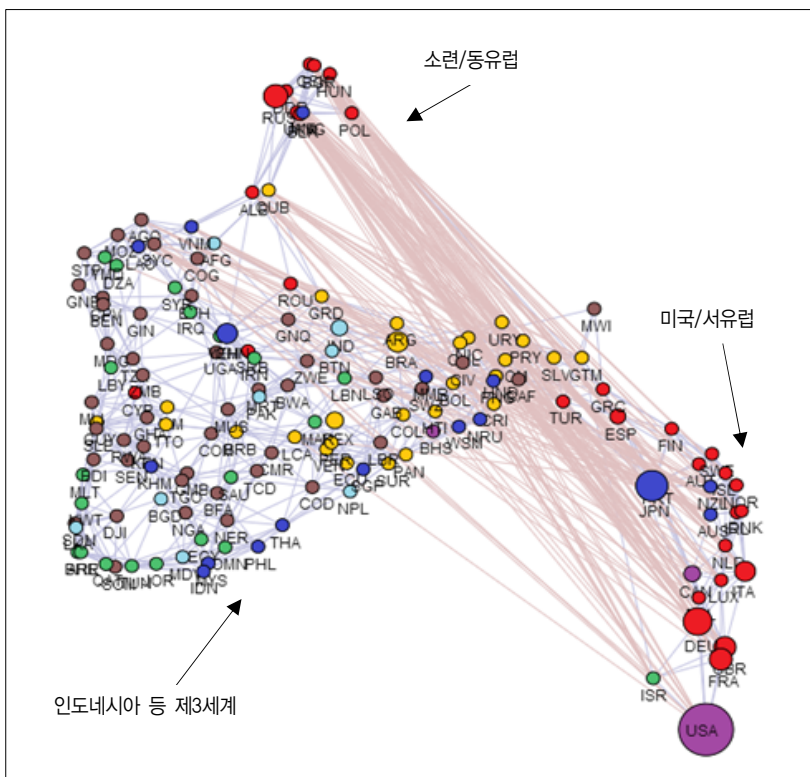
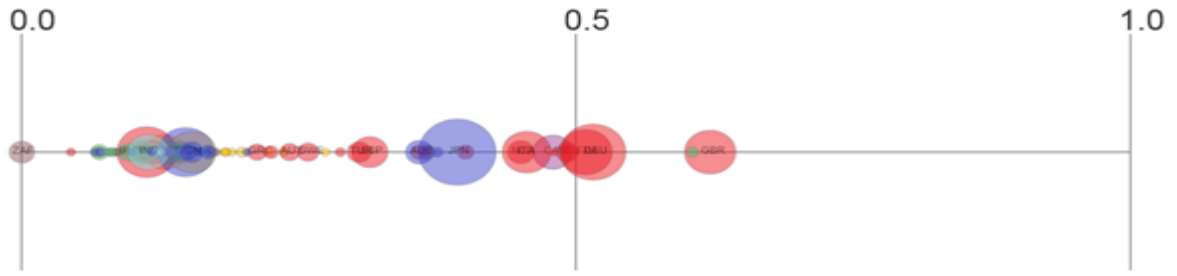
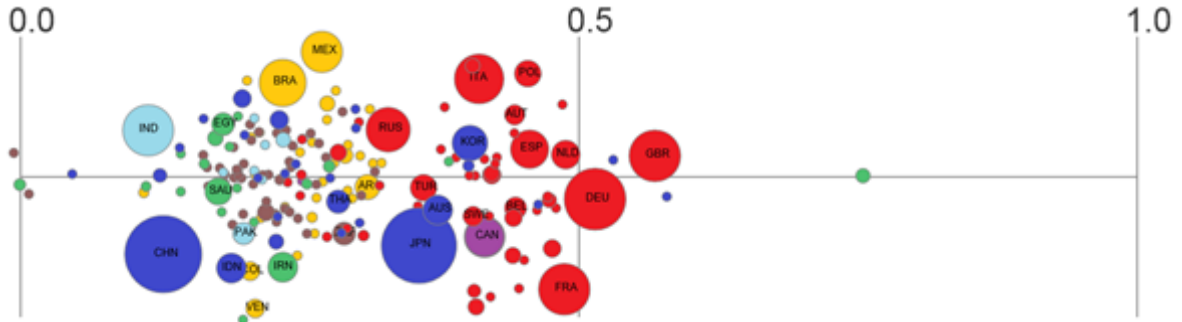


그림 8. 1976-80 UN 회원국 투표성향 (친구-라이벌 매칭 n=10)

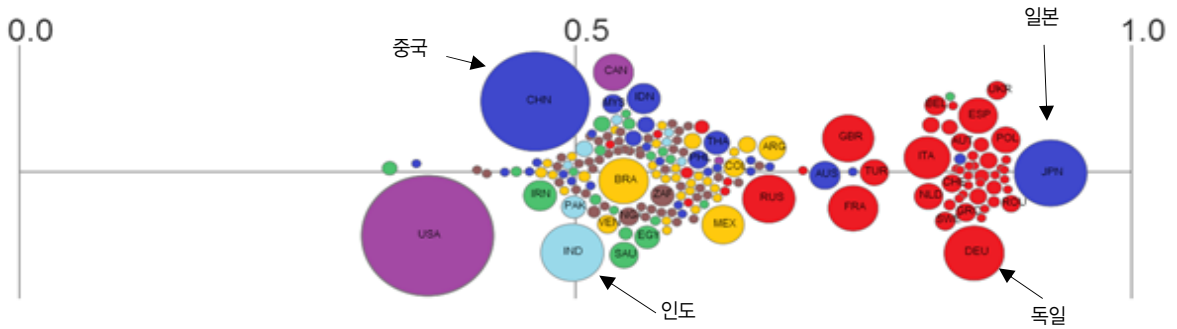


(a) 유사도 값을 x축에 배치한 회원국



(b) 버블힙 알고리즘의 중간과정

Korea, Republic of : 2011 ~ 2012



(c) 버블힙 알고리즘을 적용한 배치 결과: 우리나라를 기준으로 한 회원국의 투표성향 유사도

그림 9. 버블힙 그래프 비주얼라이제이션

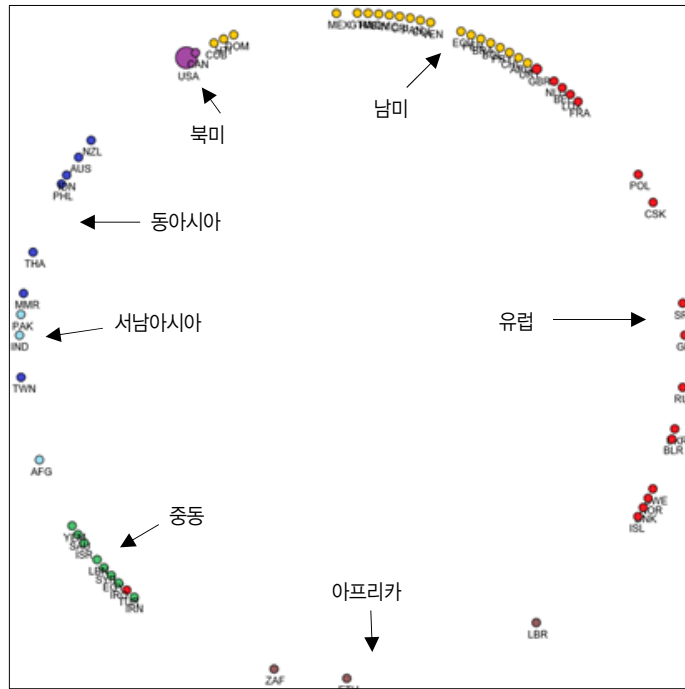


그림 10. 노드의 초기 배치

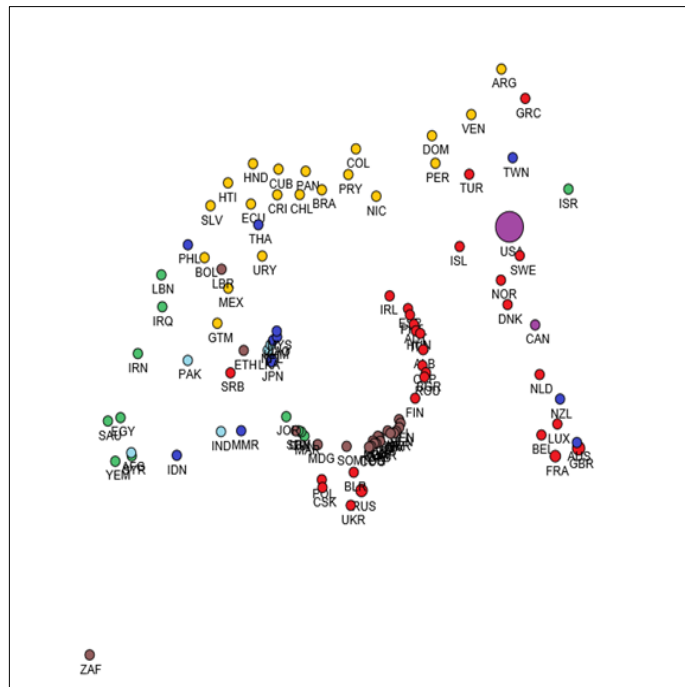


그림 11. 중간에 배치되는 회원국의 초기위치를 보여주는 그래프