creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

**저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.**

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학석사 학위논문

# 다극 시각화(N-polar Visualization)

## 다중의 앵커(anchor)를 이용한

## 인터랙티브 분석 시각화

2016 년 2 월


서울대학교 대학원

융합과학부 디지털정보융합전공

전 태 일

# N-polar Visualization

## Interactive Visual Analytics with Multiple Interactive Anchors

지도 교수 서 봉 원

이 논문을 공학석사 학위논문으로 제출함

2015 년 10 월


서울대학교 대학원

융합과학부 디지털정보융합전공

전 태 일


전태일의 공학석사 학위논문을 인준함

2015 년 12 월

위 원 장 _____ 이 원 종 _____ (인)

부위원장 _____ 서 봉 원 _____ (인)

위 원 _____ 이 준 환 _____ (인)

# 초 록

데이터 오브젝트(Data Object) 들 간에 관계(Relationship)를 이해하는 것은, 정보 시각화(Information Visualization) 분야에서 주목하고 있는 중요한 과제이며, 데이터 오브젝트 각각의 관계를 효과적으로 시각화하려는 많은 연구들이 시도되고 있다. 하지만, 일반적으로 데이터 오브젝트의 관계를 시각화한 기존의 연구들은, 전반적인 관계만을 보여주는데 그치고 있다.

본 연구에서는 정보시각화 학술분야에 존재하지 않는 새로운 시각화 테크닉인 다극 시각화(N-polar Visualization)를 제안한다. N-polar Visualization(NV)은 데이터 오브젝트간에 전반적인 관계도 볼 수 있지만, 기준이 되는 데이터 오브젝트 들을 원의 둘레에 설정할 수 있는 다중의 인터랙티브한 anchor 노드를 통해 복잡한 데이터 오브젝트들 간의 관계를 풀어낼 수 있는 것이 특징이다. 또한, 본 연구에서는 사용자의 anchor 노드에 대한 조작을 유연하게 하기위해서, anchor 노드를 원의 둘레에서 자유롭게 움직일 수 있게 하였다. 이를 통해, 사용자는 관심이 있을만한 데이터 오브젝트 들을 anchor 노드로 설정하여, 원 안쪽의 non-anchor 노드들이 선택된 anchor 노드 들과 어떤 관련성을 가지는지 살펴볼 수 있으며, 움직이는 anchor 노드에 따라 달라지는 non-anchor 노드의 패턴 변화 양상 역시 확인해 볼 수 있다.

본 연구에서는 NV 테크닉의 실행가능성(Feasibility)을 평가해보기 위해서, 데이터 오브젝트 간에 높은 관련성이 있는 특정 데이터를 적용한 일련의 예비 조사(Pilot Study)를 수행하고, 제안한 테크닉이 특정한 데이터에 관계없이 일반적(Generalization)으로 사용될 수 있는지, 추가적인 데이터셋을 가공하여 적용할 수 있었다.

마지막으로, 최종적으로 제안한 NV가 얼마나 유용(Usefulness)하고 효율적(Effectiveness)으로 사용될 수 있는지를 규명하기 위해, 패턴 연구(Pattern Study)와 사례 연구(Case Study)를 진행하였다. 첫째, UN 투표

데이터를 활용하여 어떤 종류의 패턴이 주로 발견되는지 확인하는 과정을 가지면서 다양한 패턴들을 발견할 수 있었고, 이를 2개의 그룹(Linear Pattern, Clustering Pattern)으로 분류할 수 있었다. 둘째, 실제 일어났던 역사적 사건들을 토대로 사례 연구를 진행하여 시간의 흐름에 따른 데이터에서 숨겨진 인사이트를 찾는데 있어 제안한 테크닉의 중요성을 확인해 볼 수 있었다. 또한, 발견한 인사이트와 기존에 데이터만으로는 알기 힘들었던 패턴들이 실제 발간되었던 신문기사와 일치하는 것 역시 확인할 수 있었다.

본 연구를 통해, 사용자는 데이터 오브젝트 각각에 대해서 연관성이 높은 데이터를 더 쉽게 이해할 수 있으며, 기존의 선행연구로는 알기 힘들었던 새로운 인사이트 역시 얻을 수 있을 것으로 기대한다.

**주요어 : 데이터 오브젝트, 인터랙티브 앵커, 다극 시각화, 정보 시각화**

**학 번 :** 2014-24822

# 목 차


제 1 장 서 론 ..... 1

1.1. 연구 배경 ..... 1

1.2. 연구 목적 및 프로세스 ..... 8

제 2 장 선행 연구 ..... 10

2.1. Force Directed Graph ..... 11

2.2. Multi-dimensional Scaling (MDS) ..... 13

2.3. Bubble Heap Graphs ..... 15

2.4. RadViz (Radial Visualization) ..... 16

2.5. 선행 연구 정리 및 제안 ..... 18

제 3 장 다극 시각화(N-polar Visualization) ..... 19

3.1. N-polar Visualization Mechanism ..... 19

3.2. Data Processing ..... 20

3.2.1. UN 투표 데이터(1946~2012) ..... 21

3.2.2. 미국 하원 투표 데이터(2000~2014) ..... 23

3.3. Implementation ..... 24

3.3.1. D3.js ..... 24

3.3.2. Program Structure ..... 26

3.4. Behaviors of the Nodes and the Links ..... 27

3.5. Details-on-demand and Additional Features ..... 30

3.6. Scale and Opacity ..... 31

3.7. Interactive Exploration ..... 34

**제 4 장 Pattern Study and Case Study** ..... 35
4.1. Pattern Study ..... 35
4.1.1. Linear Pattern ..... 35
4.1.2. Clustering Pattern ..... 36
4.2. Case Study ..... 37
4.2.1. 이라크 전쟁(2003~2011) ..... 37
4.2.2. 미국 대통령 대선(56회, 57회) ..... 39
4.2.3. 정리 및 한계 ..... 42
**제 5 장 결론 및 연구의 의의** ..... 43
5.1. 연구 요약 ..... 43
5.2. 연구의 시사점 ..... 44
5.3. 연구의 한계 및 제언 ..... 47
**참고문헌** ..... 49
Abstract ..... 54

# 수식 목차

[수식 1] 두 국가(Data Object)의 투표 유사도를 구하는 식 ..... 21

# 표 목차

[표 1] UN 투표 데이터 2000년의 UN Member별 유사도 결과 일부... 22

[표 2] 미국 하원 데이터 2008년의 State별 유사도 결과 일부 ..... 24

# 그림 목차

[그림 1] 정보시각화(Information Visualization)-Word Cloud ..... 1

[그림 2] 정보시각화(Information Visualization) – History Flow ..... 2

[그림 3] Flickr, Twiiter 위치 데이터를 활용한 시각화(Visualization) ..... 3

[그림 4] IBM 의 데이터 전광판(Data Wall) ..... 4

[그림 5] 한스로슬링의 갭 마인더(GapMinder) ..... 5

[그림 6] 페이스북 데이터를 이용한 소셜 네트워크 시각화 ..... 6

[그림 7] 페이스북 데이터를 이용한 Friend Wheel 시각화..... 7

[그림 8] 본 연구의 전체 프로세스..... 8

[그림 9] 선행 연구들의 장점과 한계점 요약..... 10

[그림 10] Force Directed Graph 의 예 ..... 11

[그림 11] U.S. Senate Social Graph – 미 상원의원 투표성향 시각화 ..... 12

[그림 12] Politiz – 18대 국회의원의 투표성향 시각화..... 12

[그림 13] Multi-Dimensional Scaling (MDS) 의 예 ..... 13

[그림 14] MDS 를 활용한 영화 추천 시각화..... 14

[그림 15] Bubble Heap Graphs 의 예 ..... 15
[그림 16] 유럽 에너지(Europe Energy) 시각화 (1998 ~ 2008) ..... 16
[그림 17] RadViz 시각화 테크닉의 내부 알고리즘 ..... 17
[그림 18] N-polar Visualization 과 RadViz 의 차이 ..... 19
[그림 19] D3.js 를 활용한 다양한 종류의 시각화 기법들 ..... 25
[그림 20] N-polar Visualization 프로그램 구조 ..... 26
[그림 21] N-polar Visualization - Main View 에서의 anchor 노드와 non-anchor 노드들 ..... 28
[그림 22] 1946~2012 년 까지의 UN 투표 데이터를 이용하여 시각화한 N-polar Visualization 의 5 가지 View (Main View, Control View 1, Control View 2, Detail View, Time Slider) ..... 29
[그림 23] 2000~2014년 까지의 미국 하원 투표 데이터를 이용하여 시각화한 NV의 5가지 View (Main View, Control View 1, Control View 2, Detail View, Time Slider) ..... 29
[그림 24] Scale 기능을 통한 non-anchor 노드들에 대한 구분 ..... 32
[그림 25] Opacity 기능을 통한 non-anchor 노드들에 대한 구분 ..... 33
[그림 26] Oceania 대륙 1개의 클러스터가 2가의 anchor 노드를 추가로 설정함으로써 2개의 클러스터로 나뉘어지는 사례 ..... 34
[그림 27] Linear Pattern (L1, L2, L3, L4) ..... 36
[그림 28] Clustering Pattern (C1, C2, C3, C4) ..... 36
[그림 29] 역사적 사건인 Iraq War (2003~2011) 전, 중, 후 에 대해서 UN member들 간의 관계를 밝혀낸 사례 연구 ..... 38
[그림 30] 미국의 56회(2008년), 57회(2012년) 대통령 대선에 대해서 미국의 50개 State 들 간의 관계를 밝혀낸 사례 연구 ..... 40

# 제 1 장 서 론

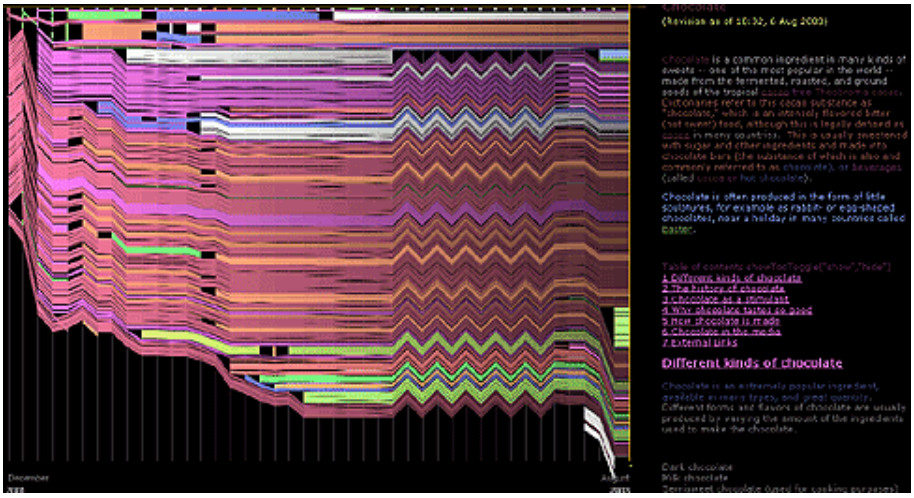
## 1.1. 연구 배경

디지털 시대의 사람들은 영상, 디지털 매체를 비롯해 SNS, 클라우드 서비스 등을 통한 무한정의 데이터에 노출되어 있으며 이와 같은 환경의 변화에 따라 데이터 표현과 수용방식의 변화가 요구되고 있다. 즉, 수 많은 정보들을 시각적으로 묘사하고 필요한 정보를 효율적이고 명확하게 제공하는 정보시각화(Information Visualization)의 필요성이 증가하고 있다.

[그림 1] : 정보시각화(Information Visualization) – Word Cloud

정보시각화는 인간의 시각 시스템에 저장된 자료들을 그대로 사용함으로써, 인간의 정보처리 능력을 향상시켜 정보를 직관적으로 이해할 수 있거나 데이터 자체로는 알지 못했던 패턴을 찾는데 도움을 준다 [1]. 또한, 시각화는 풍부한 정보를 포함하고 있어, 데이터로만 이루어져 있는

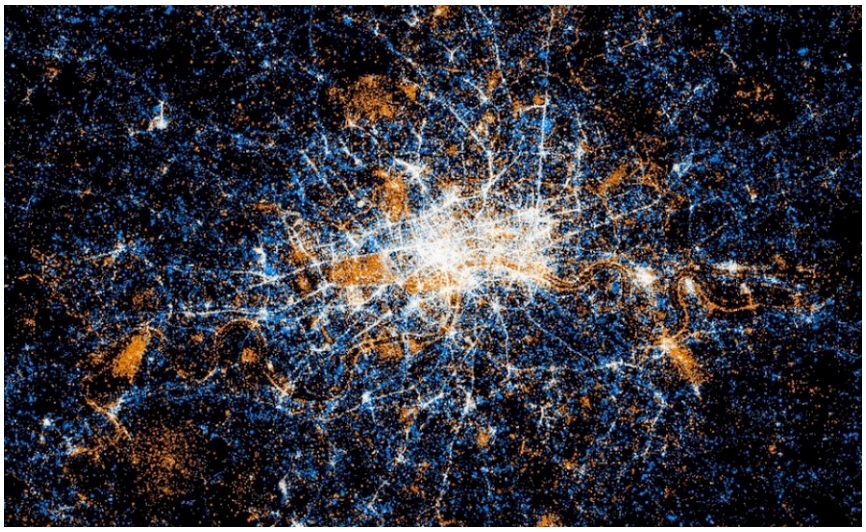
페이지보다 훨씬 더 빠르게 이해할 수 있다. 텍스트 분석의 속도가 읽기의 연속적인 과정에 의해서 제한되는 동안, 이미지 해석은 인간적인 지각 시스템 내에서 평행으로 실행되기 때문이다 [2]. 하나의 간단한 사례로 **[그림 1]** 의 단어 구름(word cloud) [3] 은 문서에 사용된 단어의 빈도를 계산해서 시각적으로 표현해주고 있다. 이 시각화는 많이 나오는 단어는 크게 표시되기 때문에 사용자는 한 눈에 문서의 핵심 내용을 파악할 수 있고, 어떤 부분이 중요하지 않은지에 대해서 단번에 알 수 있다.




[그림 2] : **정보시각화**(Information Visualization) – History Flow

정보 시각화는 그 어떤 논문보다도 훨씬 더 효과적으로 사회적 현상을 증명해보이기도 한다. 첫번째 예로, IBM 의 마틴 와텐버그(Martin Wattenberg)와 페르난도 비에가스(Fernanda Viegas)는 위키피디아(Wikipedia)에 등재되는 글이 작성되고 편집되는 과정을 시각적으로 추적하는 데이터 시각화 도구인 'History Flows' [4] 를 개발했다 **[그림 2]**. 이를 통해 위키피디아 데이터를 시각화해보면 '낙태' '이슬람' 등과 같이 첨예하게 대립하기 쉬운 주제의 글이 악의적으로 삭제될 경우

누군가에 의해 빠른 시간 내에 원상태로 복구되는 모습과 공동 저자들에 의해 위키피디아의 글이 편집되고 진화해 가는 모습이 '그랜드캐니언'의 형상처럼 표현되는 컬러 그래프의 추이를 통해 확연하게 파악할 수 있다. 두번째 사례로 구글의 Eric Fischer 는 사진 공유서비스 Flickr 와 Twitter 등 소셜 미디어 이용 현황을 지도 위에 시각화한 작품을 공개했다 [5]. 그는 위치 정보 데이터를 이용하여 Flicker 로 사진을 올리고 있는 곳은 주황색, Twitter 를 하고 있는 곳은 파란색, 둘 다 이용하고 있는 곳은 하얀색 점으로 지도 위에 표시하였다. 이런 방법을 통해서 New York, San Francisco, London 을 포함하여 세계 35 개 도시에 대한 시각화를 진행하였다. **[그림 3]** 은 영국의 London 을 시각화한 모습을 보여준다 [6]. 그 결과 많은 사람들이 35 개 도시별로 Flicker 와 Twitter 의 이용 패턴을 한 눈에 파악할 수 있었다. 또한, 도시 지도 위에서 사진이 활발하게 올라가고 있는 곳으로 표시된 지역에 대해서는 곧 관광객들이 많이 몰리는 명소일 것으로 예측이 가능하므로, 관광객은 어떤 곳이 현재 가장 인기있는 장소인지 알 수 있으며, 각국의 정부 역시 이를 활용하여 도시의 관광 정책 수립에 도움을 줄 수 있을거라 사료된다.



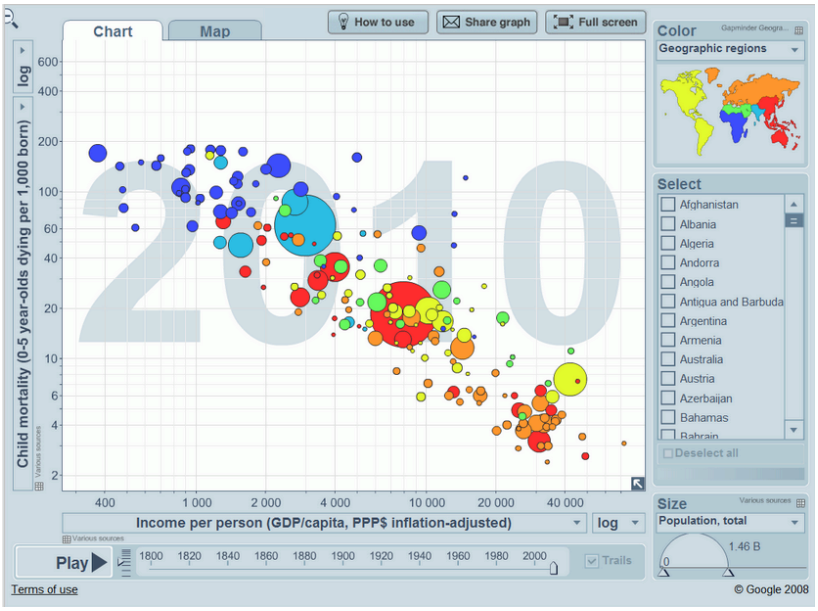
**[그림 3] : Flickr, Twiiter 위치 데이터를 활용한 시각화(Visualization)**

정보시각화는 또한 사람들의 관점을 변화시키고 세상을 이롭게 하는 행동을 촉구하기 위해 쓰이기도 한다. IBM 은 2012 년에 창립 100 주년을 기념해 뉴욕에서 'THINK' 전시회를 열고 40m 길이의 **[그림 4]**와 같은 '데이터 전광판(Data Wall)'을 선보였다 **[7]**. 전시회장 주변의 각종 데이터, 예를 들면 브로드웨이의 교통상황에서 부터 식수부족, 대기오염, 카드사고 현황 등에 이르기까지 뉴욕의 현재 삶을 느끼게 해주는 정보를 실시간으로 시각화해서 보여준다. IBM 은 이곳을 방문하는 사람들이 주변에서 어떤 일이 일어나고 있는지를 제대로 알게 하고 삶을 개선시킬 여지가 충분히 있음을 깨닫게 함으로써 세상을 지금보다 더 이롭게 만드는 것을 목적으로 이 정보 시각화를 기획했다고 한다.

**[그림 4] : IBM 의 데이터 전광판(Data Wall)**

정보시각화 분야의 각광을 받는 대표적인 연구로써 한스 로슬링의 갭 마인더(Gap Minder) **[8]** 가 있다 **[그림 5]**. 갭 마인더에서는 인구, 수출입, 환경, 교육 등 전 세계 주요 국가의 최근 200 년간의 데이터를 내장하고

있다. 재생(Play)버튼을 누르면 시간의 흐름에 따라 국가를 대표하는 원이 2 차원 평면에서 이동하게 된다. 원은 인구나 GDP 등 국가를 대표하는 값과 비례하도록 할 경우 시각적으로 해당 국가의 중요성을 인지할 수 있다. 전쟁이나 기근과 같은 특이한 경우에, 인구의 감소, 평균 수명의 변화 등 특정 국가의 이상한 움직임을 애니메이션 형태로 볼 수 있어, 정보를 한눈에 파악함과 동시에 데이터로부터 몰랐던 사실을 알게해주는 우수한 사례이다.

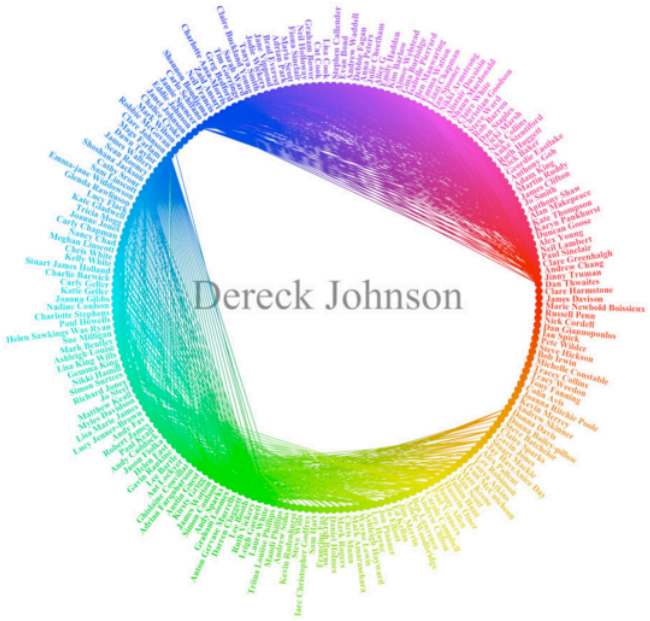



[그림 5] : 한스로슬링의 갭 마인더(GapMinder)

최근들어, 소셜 미디어의 발전으로 개인이나 조직간의 관계에 대한 관심이 많아지고 있다. 개인형 맞춤 추천시스템들의 등장과, 페이스북이나 트위터 같은 수천만명 이상의 대규모 인적 네트워크를 그래프로 구현하는 경우가 이를 대변하고 있다. [그림 6]은 페이스북 데이터를 기반으로 소셜 네트워크 시각화 [9] 를 진행한 결과물이다. 이를 통해, 자기 자신과

[그림 6] : 페이스북 데이터를 이용한 소셜 네트워크 시각화

소셜미디어(ex:페이스북)의 친구들이 전반적으로 어떠한 관계를 갖고있는지 시각적으로 한 눈에 확인 해볼 수 있다는 장점이 있다. 하지만, 대규모 인적 네트워크는 전반적인 관계나 트렌드를 파악하는데 도움이 되나 네트워크의 구성원 개개인의 성향을 파악하기에는 부족한 면이 존재한다. 이러한 소셜 네트워크 시각화의 부족한 부분을 보완하고자, 구성원 개개인의 성향을 쉽게 파악할 수 있게 진행한 시각화 작업인 Friend Wheel [10] 이라는 툴(Tool)도 있다. 이 시각화는 [그림 7]의 가운데 데이터 오브젝트(Dereck Johnson)를 기준으로 하여, 페이스북 친구들이 어떤 관계를 갖고 있는지 볼 수 있다. 또한, 학교 혹은 직장 등 비슷한 군에 속하는 집단에 대해서는 유사한 색으로 구성이되어, 기준이 되는 데이터 오브젝트의 친구들이 몇 개의 집단으로 이루어져 있는지 확인해 볼 수 도 있다. 마지막으로, 기준이 되는 데이터 오브젝트의 페이스북 친구들이 서로 동시에 아는 경우에 대해서는 링크(Link)를 이어줌으로써, 친구들 사이의 관계까지 볼 수 있는 장점을 가진다. 하지만, 이는 앞서 언급한 개개인의



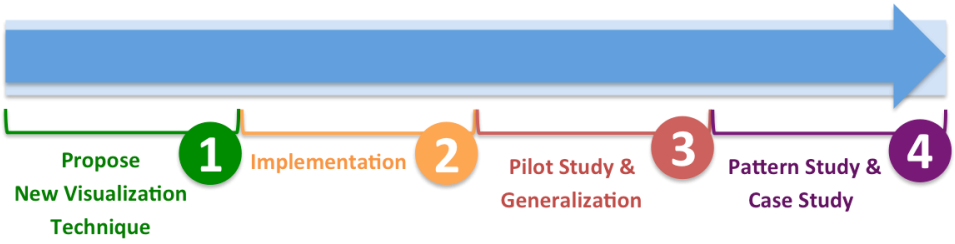

[그림 7] : 페이스북 데이터를 이용한 Friend Wheel 시각화

성향을 파악하는데 있어 유용하게 쓰일 수 있지만, 소셜 네트워크 시각화처럼 전반적인 관계나 트렌드를 보기 어렵다는 한계점이 존재한다.

이 처럼 소셜 미디어의 발전에 힘입어, 개인이나 조직간의 관계에 대해서 적절한 분석 방법이 무엇인지에 대한 논의가 산업계에서 꾸준히 이루어지고 있다 [11]. 이는 소셜 미디어 데이터에 대하여 다양한 시각화 툴이 등장하고 이에 대해 많은 연구들이 이루어지고 있음이 이를 뒷받침한다. 그리고, 산업계 뿐만 아니라, 데이터 오브젝트 (개인 or 조직) 들 간에 관계(Relationship)를 이해하는 것은, 정보시각화 (Information Visualization) 학술 분야에서도 최근 주목하고 있는 중요한 과제이며 [12], 데이터 오브젝트 각각의 관계를 효과적으로 시각화하려는 많은 연구들이 시도되고 있다.

## 1.2. 연구 목적 및 프로세스

본 연구에서는 데이터 오브젝트(Data Object) 각각의 관계를 시각화한 선행 연구들의 테크닉을 분석하여, 기존 테크닉들이 가지고 있는 한계점을 개선하고, 시각화 분야에서 존재하지 않는 새로운 테크닉을 제안하는 것을 그 목적으로 한다. 정보 시각화 분야에서, 데이터 오브젝트 각각의 관계를 효과적으로 시각화려는 선행 연구들이 있었지만, 대체적으로 데이터 오브젝트들 끼리의 전반적인 관계만을 보여주는데 국한 되어 왔기 때문이다. 이에 본 연구는 [그림 8] 과 같이, 크게 4 가지의 프로세스로 진행된다.




[그림 8] : 본 연구의 전체 프로세스

(1) 기존의 데이터 오브젝트 관계를 효과적으로 시각화한 여러 테크닉들을 분석하여, 그들이 가지고 있는 한계점을 개선할 수 있는, 새로운 시각화 테크닉을 제안(Propose)한다.

(2) 새로운 테크닉의 내부 알고리즘과 사람들이 사용할 수 있는 프로그램을 구현(Implementation)하고, 추가적으로 필요한 인터랙션 기능들에 대한 개발 작업을 진행한다.

(3) 제안한 테크닉의 실행가능성(Feasibility)을 평가해보기 위해서, 데이터 오브젝트 간에 높은 관련성이 있는 특정 데이터를 적용한 일련의 예비 조사(Pilot Study)를 수행한다. 그 다음, 예비 조사를 진행하면서 발생한 시각화 테크닉의 부족한 부분을 수정하고, 제안한 테크닉이 데이터의 특정

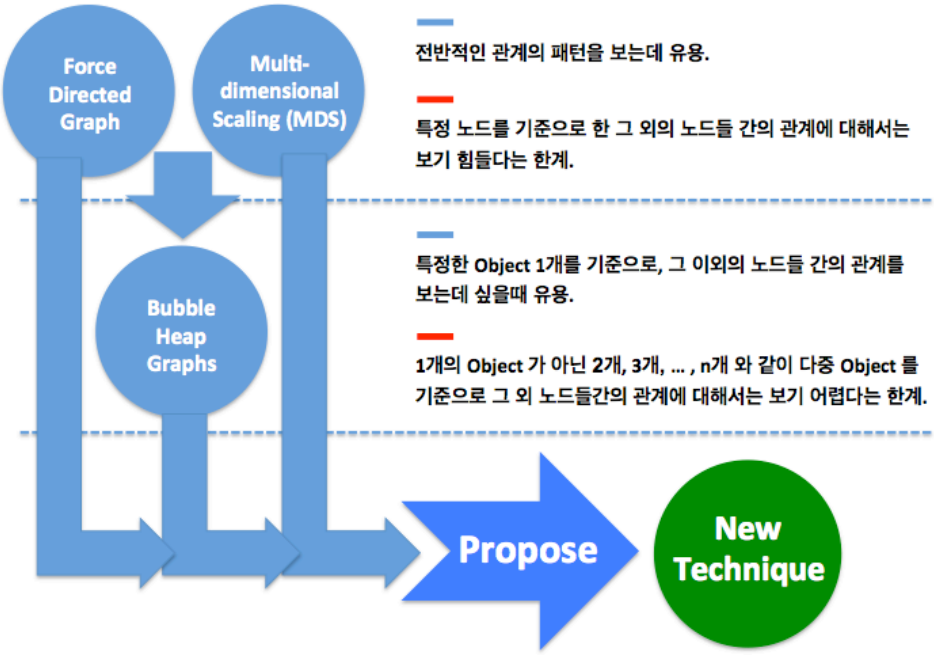
분야에 관계없이 일반적(Generalization)으로 사용될 수 있는지, 추가적인 데이터셋을 가공하여 적용해 본다.

(4) 제안한 테크닉의 유용성(Usefulness)과 효율성(Effectiveness)을 검증해보기 위해서 두 가지 데이터셋을 활용한 패턴 연구(Pattern Study) 와 사례 연구(Case Study)를 진행한다.

같은 데이터라도 보여주는 시각화 방식을 바꿀 경우에 사용자는 데이터를 더 쉽게 이해하고 새로운 인사이트를 얻을 수 있다 [13]. 본 연구를 통해, 사용자는 데이터 오브젝트 각각에 대해서 연관성이 높은 데이터를 더 쉽게 이해하고, 기존의 선행연구들이 수행하기 어려웠던 태스크(Task)들을 시도해봄으로써, 기존 방법으로는 보지 못했던 새로운 인사이트 역시 얻을 수 있을 것으로 기대된다.

# 제 2 장 선행 연구

본 장에서는 데이터 오브젝트(Data Object) 각각의 관계를 효과적으로 시각화하려는 선행 연구들을 정리하고, 그들이 가지는 장점과 한계점, 마지막으로 어떤 분야의 연구에서 해당 시각화 테크닉들이 사용 되었는지를 서술한다 [그림 9]. 그리고, 기존의 선행 연구들이 가지는 한계점을 개선할 수 있을만한 메카니즘(Mechanism)을 활용하여 기존에 존재하지 않는 새로운 시각화 테크닉을 제안하려고 한다.

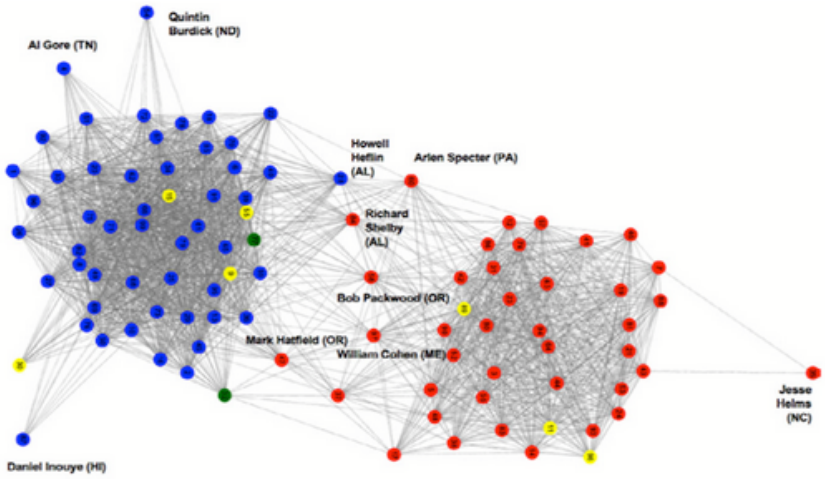



[그림 9] : 선행 연구들의 장점과 한계점 요약

## 2.1. Force Directed Graph

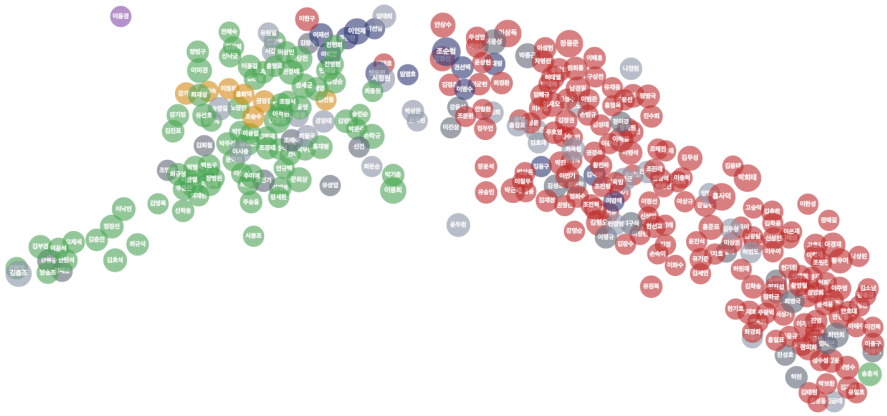
[그림 10] : Force Directed Graph 의 예

Force Directed Graph 의 내부 알고리즘은, 평형점을 찾는 것을 목표로 하는 대신 링크(Link)가 있는 노드(Data Object) 간에는 인력이 작용하며 링크가 없는 노드 간에는 척력이 작용하는 모델을 사용한다 [14]. 처음에는 노드의 위치를 랜덤으로 정하고 노드에 가해지는 외력에 따라 새로운 위치를 계산한다. 이러한 사이클을 반복하면서 외력에 의해 이동할 수 있는 거리를 점차 축소하여 안정된 위치를 찾는 방식이다. [그림 10] 과 같이, 이 시각화 방법은 노드(Data Object) 들 간의 전반적인 관계를 한 눈에 보여주는데 널리 사용되며, 가까운 관계에 있는 노드들끼리는 모이는 경향이 생기고, 먼 관계에 있으면 멀리 떨어지는 경향이 생긴다. [그림 10] 을 보면, 파란색과 하늘색 노드들의 경우 다른색을 가졌지만 근처에 위치하고 있어 비교적 높은 관계에 있다고 볼 수 있고, 파란색과 보라색 노드는 거리가 떨어져있어, 이전꺼에 비해 상대적으로 낮은 관계에 있다고 말할 수 있다. 이러한 Force Directed Graph 알고리즘의 장점인 전반적인 관계에 대한 패턴을 볼 수 있다는 점을 이용한 시각화 연구 몇 개를 소개한다.




[그림 11] : U.S. Senate Social Graph – 미 상원의원 투표성향 시각화

첫째로, 오데반은 미 상원의원의 투표 성향을 1991 년 부터 2008 년까지 분석하여 유사성향의 의원의 연결망을 Force Directed Graph 알고리즘을 이용해 [그림 11] 과 같이 시각화를 하였다 [15]. 확고한 양당체제가 자리잡고 있는 미국의 경우 민주당과 공화당의 의원으로 구성된 두개의 클러스터가 존재하며 중간에 몇 명의 의원이 중간자 역할을 하고 있음을 보여준다.

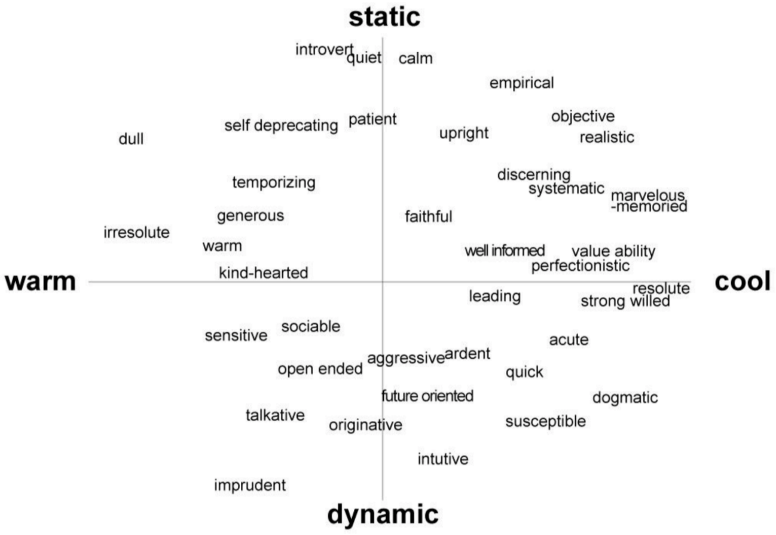


[그림 12] : Politiz – 18 대 국회의원의 투표성향 시각화

둘째로, 18 대 국회의원의 투표성향을 시각화 한 Politiz 가 있다 [16]. 해당 시각화의 경우, 의원 노드간의 링크를 Force Directed Graph 를 이용해서 의원들의 전반적인 관계를 쉽게 이해할 수 있도록 시각화를 구성하였다 [그림 12]. 이 역시 새누리당과 민주당의 의원들로 구성된 두개의 클러스터가 존재하며, 몇 개의 가외치(Outlier) 노드들이 보임을 알 수 있다.

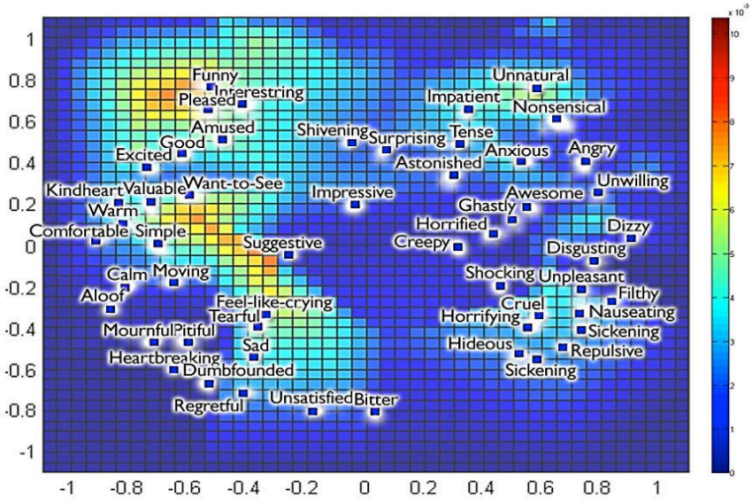
하지만, Force Directed Graph 시각화는 전반적인 패턴만을 보는데 유용하고, 특정 노드를 기준으로 한 그 외의 노드들 간의 관계에 대해서는 보기 힘들다는 어려움이 존재한다.

## 2.2. Multi-Dimensional Scaling (MDS)




[그림 13] : Multi-Dimensional Scaling (MDS) 의 예

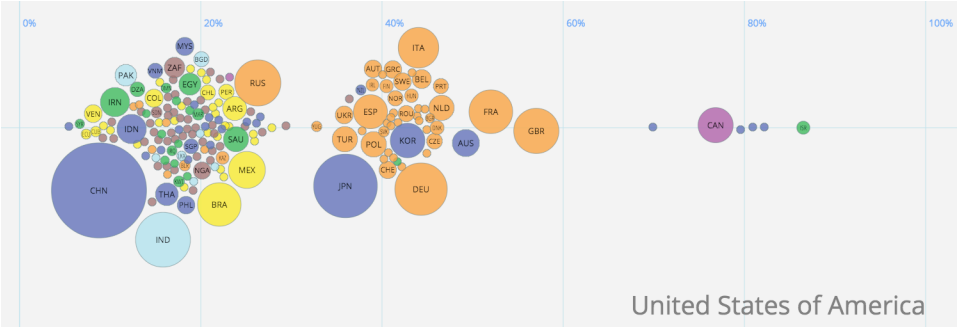
Multi-Dimensional Scaling (MDS) 은 데이터 오브젝트(Data Object)들 간의 상대적인 거리를 계산하여 사람이 인지 할 수 있는 평면상에 상대적인 거리로 나타내는 방법이다 [17]. 이는 통계분석과 정보시각화에서 유사성과 차별성을 살펴보기 위한 일반적인 차원 감소 알고리즘의 대표적 종류라고 할 수 있다. MDS 의 장점은 상대적인 거리만을 알고 있는 데이터 오브젝트들의 위치도를 작성할 수 있으며 물리적인 거리뿐만 아니라 심리적인 거리에 근거하여서도 위치도를 작성할 수 있다는 점이다. MDS 의 경우 데이터 오브젝트들 사이의 전반적인 관계를 보고자 시각화 분야에서도 많이 사용되고 있으며, 특히, **[그림 13]** 에서 처럼 단어(Data Object)와 단어(Data Object)사이의 의미상의 거리를 2 차원 평면에 뿌리고자 할 때, 자주 사용되고 있다. **[그림 14]** 는 MDS 를 활용한 정보시각화 연구이다 [18]. 이 시각화는, 영화 50 개를 가지고, 100 명 가량의 사람들이 영화를 보고 느끼는 감정상태를 어휘로 만든 뒤, 이를 MDS 시각화를 통해서 의미상 가까운 어휘(Data Object)들은 붙어있고, 의미상 먼 어휘(Data Object)들은 떨어져 있는 시각화 뷰를 제공하였다. 해당 테크닉을 통해서, 사람들은 자신의 현재 감정상태에 따른 영화를 감정어휘의 전반적인 뷰를 보면서 추천받을 수 있다.




**[그림 14] : MDS 를 활용한 영화 추천 시각화**

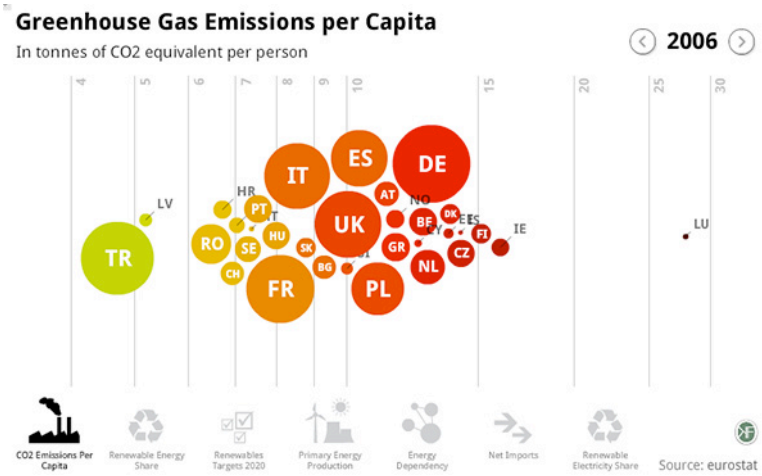
하지만, 이 MDS 시각화 역시, 이전의 Force Directed Graph 와 마찬가지로, 전반적인 패턴만을 보는데 유용하고, 특정 노드(Data Object)를 기준으로 한 그 외의 노드(Data Object)들 간의 관계에 대해서는 보기 힘든 어려움이 있다.

## 2.3. Bubble Heap Graphs




[그림 15] : Bubble Heap Graphs 의 예

앞서 언급한 2 개의 전반적인 관계만 보여줄 수 있는 한계를 극복해보려고 시도된 시각화 연구 중에 Bubble Heap Graphs 가 있다 [19]. Bubble Heap Graphs 는 사용자가 특정한 데이터 오브젝트(Data Object) 1 개를 기준으로, 그 이외의 노드(Data Object)들 간의 관계를 알고 싶을때 유용한 시각화이다. **[그림 15]** 를 보면, 미국을 기준으로 각 국가별 유사도 정도를 왼쪽 0% ~ 오른쪽 100% 로 설정하여 미국과 관계가 있는 국가들의 상대적인 거리를 일렬로 시각화 해주고 있다. 왼쪽에 위치한 노드들은 미국과의 유사도가 낮은것이며, 오른쪽에 배치될 수록 높은 유사도를 가진다. ( **[그림 15]** 오른쪽 밑에 United States of America 가 기준이 되는 노드(Data Object)이다. )



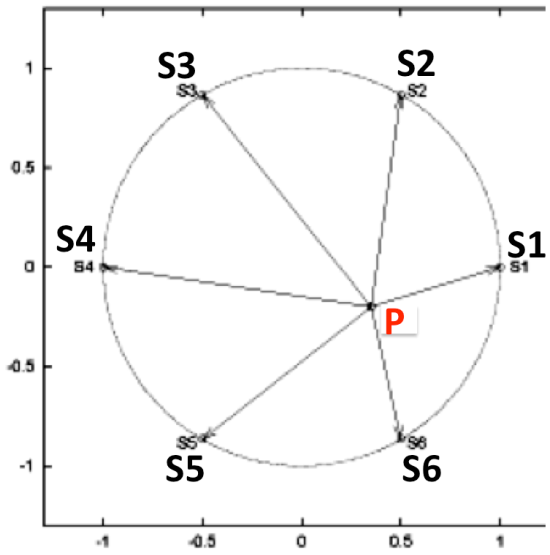

[그림 16] : 유럽 에너지(Europe Energy) 시각화 (1998 ~ 2008)

해당 시각화를 활용한 예로, [그림 16] 의 유럽 에너지(Europe Energy) [20] 라는 비쥬얼라이제이션이 있다. 이것은 유럽국가의 에너지 생산, 이산화탄소 배출, 재생 에너지 등 에너지 정책과 관련된 데이터를 애니메이션을 통해 보여준다. 왼쪽에는 낮은 유사도, 오른쪽은 높은 유사도로 설정하여 유럽국가들을 Bubble 로 표현하여 왼쪽부터 오른쪽 일렬로 시각화 해주고 있다.

하지만, Bubble Heap Graphs 의 경우 사용자가 1 개의 데이터 오브젝트가 아닌 2 개, 3 개, … , n 개 와 같이 다중 데이터 오브젝트를 기준으로 그 외 노드들간의 관계에 대해서는 보기 어렵다는 한계가 있다.

## 2.4. RadViz (Radial Visualization)

본 절에서는, 이전 선행연구들의 한계점을 개선 하기 위한 본 연구와 비슷한 연구라 할 수 있는 RadViz 를 설명한다.




[그림 17] : RadViz 시각화 테크닉의 내부 알고리즘

RadViz(Radial Visualization)는 각각의 Dimension 을 원 둘레 주위의 anchor(Attribute) 로 배치시킴으로써, Multi-dimensional dataset 을 시각화 하는데 있어서 유용한 방법이다 [21]. [그림 17] 을 보면, 원 둘레에 s1, s2, s3, s4, s5, s6 의 Attribute 가 있고, 한 점 P (i.e : Data Object) 가 있다. P 는 원 둘레에 있는 Attribute(anchor) 들이 당기는 힘에 따라 안쪽의 어느 지점에 배치되는 지가 결정된다. [그림 17] 에서 P 점은, (s1, s6) 2 개의 Attribute 에 대해서 가까이 배치되어 있고, 나머지 (s2, s3, s4, s5) 의 Attribute 하고는 멀리 떨어져 있음을 볼 수 있다. 이는, (s1, s6) 가 P 점을 강하게 당기고 있다는 것을 의미하며, (s2, s3, s4, s5) 는 P 점을 약하게 당기고 있다는 것을 의미한다. 풀어서 설명하자면, P 점은 (s1, s6) Attribute 와는 높은 관련이 있는 반면에, (s2, s3, s4, s5) 하고는 낮은 관련이 있다는 것과 일맥 상통한다. 이러한, RadViz 시각화 방법을 통해서, 사용자는 Data Object 들이 어느 쪽으로 편중되어 있고 밀집되어 있는지 확인해 볼 수 있다.

이처럼, RadViz 시각화는 노드들 간의 전반적인 관계를 보여주는과 동시에, 기준이 되는 Attribute 들을 원의 둘레에 설정함으로써 안 쪽 노드들과 바깥 쪽 노드들 간의 다중 비교를 가능하게 한 점이 특징이다. 본 연구에서는 이러한 RadViz 메카니즘을 활용하여, 새로운 시각화 테크닉을 제안하려고 한다.

## 2.5. 선행 연구 정리 및 제안

데이터 오브젝트의 관계를 시각화한 기존의 선행연구들은, 전반적인 관계만을 보여주는데 그치고 있다. 또한, 전반적인 관계만을 보여줄 수 있는 한계를 극복해보려고 시도된 Bubble Heap Graphs 역시 특정 데이터 오브젝트 1개를 기준으로, 그 이외의 노드(Data Object)간의 관계만을 보는데만 유용하게 사용되며, 다중의 데이터 오브젝트를 기준으로 그 외 노드들 간의 관계에 대해서는 보기 어렵다는 한계점이 존재한다. 이에 본 연구에서는, 이러한 선행 연구 들이 가지는 한계점들을 극복하기 위해 RadViz 메카니즘을 일부 활용한 새로운 시각화 테크닉을 제안하였다.

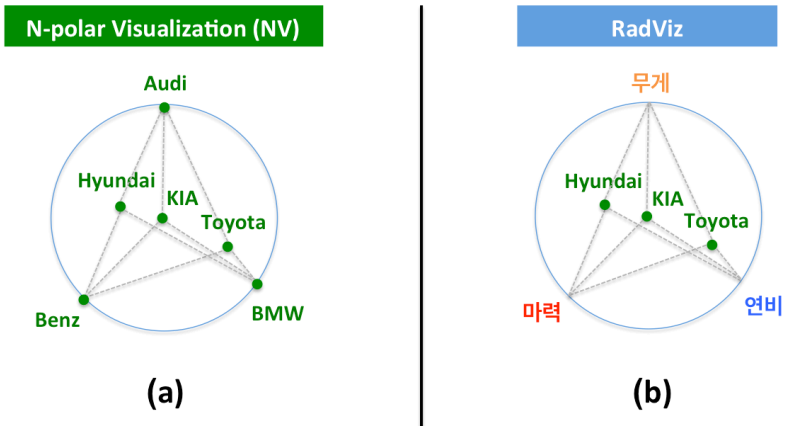
본 시각화 연구를 통해서, 사용자는 데이터 오브젝트 간의 전반적인 관계를 볼 수 있을 뿐만 아니라, 다중의 데이터 오브젝트를 기준으로한 그 외 노드들 간의 관계 역시 확인해볼 수 있다. 또한, 데이터 오브젝트 각각에 대해서 관계가 복잡한 데이터를 가지고 있을 때, 데이터 오브젝트들 간의 관계를 쉽게 풀어낼 수 있어, 가진 데이터를 더 쉽게 이해할 수 있다. 마지막으로, 기존의 선행연구들이 수행하기 힘들었던 태스크(Task) 역시 시도해볼 수 있어 기존의 시각화 방법으로는 알지 못했던 새로운 인사이트 역시 얻을 수 있을것으로 기대한다.

# 제 3 장 다극 시각화(N-polar Visualization)

본 장에서는, RadViz 메카니즘을 활용한 기존에 존재하지 않는 새로운 테크닉인 N-polar Visualization(NV) 을 제안하며 장의 구성은 다음과 같다. 첫째, RadViz 와 NV 의 차이점을 통해, 어떤 부분이 비슷하고 차별화 되는지에 대해서 알아본다. 둘째, NV 테크닉의 실행가능성(Feasibility) 과 이를 일반화(Generalization)하기 위한 두 종류의 데이터셋 소개, 그리고 어떤 방법으로 데이터를 가공했는지 설명한다. 셋째, NV 를 구현하는데 사용한 라이브러리를 소개하고, 전반적인 프로그램 구조에 대해 서술한다. 마지막으로, NV 의 사용자 인터페이스와 인터랙션 요소, 추가적으로 구현한 기능들에 대해 소개한다.

## 3.1. N-polar Visualization Mechanism

본 연구에서 N-polar Visualization 은 RadViz 시각화의 메카니즘인 각각의 Dimension 들을 원 둘레의 anchor 들로 배치시키는 방법을 활용하여, 선행 연구들이 가지고 있었던 한계점을 개선 한다.




[그림 18] : N-polar Visualization 과 RadViz 의 차이

RadViz 와 N-polar Visualization(NV) 의 주요한 차이는, RadViz 는 원 둘레에 Attribute 가 배치가 되고, 데이터 오브젝트(Data Object)가 원의 안에 배치가 된다는 점이다. 반대로, NV 는 데이터 오브젝트가 원의 둘레와, 안쪽에 모두 배치가 되는 차이가 있다. 예를 들어서 설명하자면, N-polar Visualization 은 사용자가 Audi, Benz, BMW 를 관심으로 두었을때, 설정한 3 개의 브랜드(Data Object)와 그 외의 브랜드(Data Object)들과의 상대적인 위치를 본다. **[그림 18]** 의 **(a)**를 보면 Hyundai 노드의 경우 Benz, Audi 와는 상대적으로 가까운 반면에 BMW 하고는 멀다. KIA 노드의 경우는 3 개의 기준 데이터 오브젝트와의 상대적인 거리가 유사하다. 마지막으로, Toyota 노드는 BMW 하고 상대적으로 가깝지만, Audi, Benz 하고는 멀다.

반대로, RadViz 의 경우는 브랜드(Data Object) 끼리의 상대적인 위치를 보는 것이 아니라, 외부에 데이터 오브젝트가 아닌, 다른 요소들(Attribute)이 지정되는 개념이다. **[그림 18]** 의 **(b)**를 보면, 원의 둘레에 무게, 마력, 연비와 같이 기존의 브랜드(Data Object)가 아닌, 다른 요소(Attribute) 가 들어감을 알 수 있다.

그래서 NV 는, 데이터 오브젝트를 원 둘레의 anchor 로 배치시키는 방법을 활용하여 데이터 오브젝트들간의 복잡한 관계를 풀어낼 수 있다는 장점을 가진다. 다시말해서, 전체 데이터 오브젝트의 일부를 외부의 원 둘레에 두게되면, 내부의 데이터 오브젝트들은 자연스럽게 감소하게 됨으로써 복잡하게 엮여있는 내부의 관계가 풀린다는 개념이다.

## 3.2. Data Processing

본 절에서는 N-polar Visualization 테크닉의 데이터로 사용된 데이터 오브젝트(Data Object)들 끼리 높은 연관성을 가지는 두 종류의 서로 다른 데이터셋을 설명한다. 또한, 각각의 데이터셋이 어떤 형태로 구성되어

있으며, 이를 어떤 방식으로 가공하고 예외처리 하여 최종 시각화에 적용하였는지 서술한다.

### 3.2.1. UN 투표 데이터 (1946~2012)

본 연구에서는 첫번째 데이터셋으로 1946~2012 년 까지의 UN 투표 데이터 [22] 를 이용하여, N-polar Visualization 을 시각화하는 데에 적용해 보고자 했다. UN 투표 데이터는, 각각의 UN 에 가입한 국가들이 각각의 issue 에 대해서 "Yes", "Abstain", "No", "Absent" 의 의사표시가 쓰여 있으며, 해당 시기에 UN 에 가입되지 않은 국가들은 "Not a UN Member" 가 기록되어 있다. 본 연구에서는 데이터 오브젝트(UN Member)들 끼리의 관계를 보기 위해, 투표 유사도 값을 계산하였다. 계산식은, 2 개의 UN Member 를 기준으로 투표 결과가 얼마나 유사한지에 대해서 Similarity(x ,y) **[수식 1]**을 따른다.

$$Similarity(x, y) = \frac{\text{the number of issues where both nations agree}}{\text{the total number of issues}}$$

**[수식 1] : 두 국가(Data Object)의 투표 유사도를 구하는 식**

Similarity(x, y) 는 2 개의 국가 x, y 에 대해서 투표 결과의 유사한 정도를 나타낸다. Similarity(x, y) 값이 1 에 가까울 수록 x 와 y 는 투표 결과가 유사하다는 것을 의미하고, 0 에 가까울 수록 유사하지 않다는 것을 의미한다. 이러한 계산 방법으로, 모든 UN Member 들에 대해서 1946 년부터 2012 년까지 1 년을 단위로 유사도(Similarity)값을 계산하였다. 그러나,

x 혹은 y 값 중 하나에 “Not a UN Member” 가 기록되어 있다면, 해당 Similarity(x, y) 는 계산하지 않았다.

본 연구에서 UN Member 를 표시할 때 2 자 또는 3 자의 알파벳을 사용하였다. 국제 올림픽 위원회 (IOC), 국제축구연맹(FIFA) 와 같이 전세계의 국가를 대상으로 하는 기관은 나름대로 독자적인 국가코드를 사용하고 있다. 본 논문에서 국가코드는 국제표준기구(ISO) 의 ISO 3166-1 alpha-3 을 사용하였다 [23]. UN 투표 데이터는 총 198 개의 회원국 수를 가지며 67 년이라는 기간 동안 국가에 대한 변화 사항들이 있어 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 멘탈모델에 맞는 합리적인 처리가 필요하다. 본 연구에서 독일의 경우에는 East Germany 가 West Germany 로 흡수통합으로 West Germany 의 데이터를 Germany 로 함께 사용하였으며, Czechoslovakia 의 경우에는 Czech 와 Slovakia 의 두 국가로 분리 표시되며 Czechoslovakia 는 별도의 국가로 정의하였다. 소련연방 해체의 경우에는 Russia 가 소련을 이어받도록 처리하였으며, Yugoslavia 의 경우에는 Serbia 가 Yugoslavia 를 계승하는 것으로 처리하였다. 마지막으로 Zanzibar 는 Raw 데이터의 유실이 있어 제외 국가로 선정하였다.

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2000 | USA | CAN | BHS | CUB | HTI | DOM | JAM | TTO | BRB | DMA |
| 2 | USA | -2 | 0.3864 | 0.1879 | 0.092 | 0.203 | 0.2179 | 0.1744 | 0.1849 | 0.1825 | 0.2055 |
| 3 | CAN | 0.3864 | -2 | 0.5987 | 0.4632 | 0.5664 | 0.5783 | 0.5773 | 0.5836 | 0.5901 | 0.5324 |
| 4 | BHS | 0.1879 | 0.5987 | -2 | 0.7807 | 0.8529 | 0.8781 | 0.8982 | 0.8881 | 0.9272 | 0.8956 |
| 5 | CUB | 0.092 | 0.4632 | 0.7807 | -2 | 0.7706 | 0.7346 | 0.8054 | 0.7869 | 0.7856 | 0.7862 |
| 6 | HTI | 0.203 | 0.5664 | 0.8529 | 0.7706 | -2 | 0.874 | 0.8498 | 0.8408 | 0.8545 | 0.851 |
| 7 | DOM | 0.2179 | 0.5783 | 0.8781 | 0.7346 | 0.874 | -2 | 0.8463 | 0.8473 | 0.8633 | 0.8467 |
| 8 | JAM | 0.1744 | 0.5773 | 0.8982 | 0.8054 | 0.8498 | 0.8463 | -2 | 0.8875 | 0.8979 | 0.9126 |
| 9 | TTO | 0.1849 | 0.5836 | 0.8881 | 0.7869 | 0.8408 | 0.8473 | 0.8875 | -2 | 0.8996 | 0.9068 |
| 10 | BRB | 0.1825 | 0.5901 | 0.9272 | 0.7856 | 0.8545 | 0.8633 | 0.8979 | 0.8996 | -2 | 0.9092 |
| 11 | DMA | 0.2055 | 0.5324 | 0.8956 | 0.7862 | 0.851 | 0.8467 | 0.9126 | 0.9068 | 0.9092 | -2 |

[표 1] : UN 투표 데이터 2000 년의 UN Member 별 유사도 결과 일부

[표 1] 은 위의 데이터 가공 방법을 토대로 UN 투표 데이터 2000 년의 UN Member 들 각각에 대한 유사도를 계산한 매트릭스 중 일부를 보여준다. 여기서 USA(미국) 과 CAN(캐나다) 의 경우 2000 년에 유사도가 0.3864, USA(미국) 과 CUB(쿠바) 의 경우는 유사도가 0.092 인 것을 알 수 있다. 대각선으로 '-2' 라고 쓰여있는 값은, 'NA' 대신으로 넣어 둔 값을 의미하는데 특정 시기에 UN 에 가입되지 않은 UN Member 는 모두 '-2' 값이 들어간다.

### 3.2.2. 미국 하원 투표 데이터 (2000~2014)

두 번째 데이터셋으로는 미국 하원(House) 의원의 2000~2014 년 동안에 "Amendment" 카테고리의 법안에 대한 투표 데이터 [24] 를 사용하였다. 해당 하원 투표 데이터는 Govtrack.org 에서 데이터베이스 형태로 일반에 공개하고 있으며, 미국의 50 개 State 각각에 소속되어 있는 하원 의원 마다 각각의 법안에 대해서 "Yea", "Nay", "No Vote" 3 가지의 의사표시가 쓰여 있다.

본 연구에서는 UN 투표데이터와 마찬가지로 데이터 오브젝트(U.S. State)들 끼리의 관계를 보기위해, 투표 유사도 값을 계산하였다. 이전 UN 투표 데이터 에서는 2 개의 UN Member 에 대해서 계산했다면, 미국의 하원 데이터는 2 개의 State 를 기준으로 투표 결과가 얼마나 유사한지에 대해서 계산하였다. 계산식은 UN 투표 데이터와 비슷한 맥락으로, State A 와 B 간의 투표 유사도는 두 State 의 하원의원들이 같은 의견을 표시한 투표 횟수를 두 State 의 하원의원들이 동시에 투표에 참여한 횟수로 나눈 값으로 정의한다. 해당 데이터의 State 를 표시하는 경우 2 자의 알파벳을 사용하였으며, UN 투표 데이터에 속해있는 UN Member 를 표시할때와 마찬가지로 국제표준기구(ISO) 의 ISO 3166-1 alpha-3 을 사용하였다 [23].

| | A | B | C | D | E | F | G | H | I | J | K |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2008 | AL | AK | AZ | AR | CA | CO | CT | DE | FL | GA |
| 2 | AL | -2 | 0.8734 | 0.7243 | 0.9542 | 0.1535 | 0.4952 | 0.0524 | 0.0433 | 0.4554 | 0.6564 |
| 3 | AK | 0.8734 | -2 | 0.7945 | 0.8183 | 0.2942 | 0.4824 | 0.1264 | 0.1125 | 0.6235 | 0.7977 |
| 4 | AZ | 0.7243 | 0.7945 | -2 | 0.6934 | 0.4893 | 0.5733 | 0.2312 | 0.2252 | 0.7923 | 0.9542 |
| 5 | AR | 0.9542 | 0.8183 | 0.6934 | -2 | 0.1285 | 0.4732 | 0.0458 | 0.0323 | 0.4925 | 0.6236 |
| 6 | CA | 0.1535 | 0.2942 | 0.4893 | 0.1285 | -2 | 0.8212 | 0.8552 | 0.8768 | 0.635 | 0.4743 |
| 7 | CO | 0.4952 | 0.4824 | 0.5733 | 0.4732 | 0.8212 | -2 | 0.7923 | 0.8023 | 0.8923 | 0.6424 |
| 8 | CT | 0.0524 | 0.1264 | 0.2312 | 0.0458 | 0.8552 | 0.7923 | -2 | 0.9525 | 0.4352 | 0.2534 |
| 9 | DE | 0.0433 | 0.1125 | 0.2252 | 0.0323 | 0.8768 | 0.8023 | 0.9525 | -2 | 0.4142 | 0.2412 |
| 10 | FL | 0.4554 | 0.6235 | 0.7923 | 0.4925 | 0.635 | 0.8923 | 0.4352 | 0.4142 | -2 | 0.7992 |
| 11 | GA | 0.6564 | 0.7977 | 0.9542 | 0.6236 | 0.4743 | 0.6424 | 0.2534 | 0.2412 | 0.7992 | -2 |

[표 2] : 미국 하원 데이터 2008 년의 State 별 유사도 결과 일부

[표 2] 는 미국 하원 투표 데이터 2008 년의 State 들 각각에 대한 유사도를 계산한 매트릭스 중 일부를 보여준다. 여기서 AL(Alabama) State 와 AK(Alaska) State 의 경우 2000 년에 유사도가 0.8734 임을 알 수 있으며, UN 투표 데이터와 마찬가지로, '-2' 라고 쓰여있는 경우는 NA 대신으로 넣어 둔 값을 의미한다.

## 3.3. Implementation

본 절에서는 N-polar Visualization(NV) 을 구현하는데 중점적으로 사용된 D3.js 에 대해 설명하고, NV 프로그램 구조에 대한 상세 내용을 서술한다.

### 3.3.1. D3.js

N-polar Visualization(NV) 을 구현하기 위한 프로그램은 모든 브라우저에서 지원하는 HTML5 를 기반으로 한다. 따라서 프로그램 언어로는 HTML 의 기본 프로그램 언어인 자바스크립트(Javascript)를

사용하였다. 데이터 분석을 통해 얻어진 결과 값을 시각화하기 위해서는 여러 기술을 사용할 수 있다. 현재 웹(Web) 상에서 가장 많이 사용하고 있는 것은 Google 에서 지원하는 Google Chart API [25] 와, HTML5 + 자바스크립트 기반의 D3.js [26] 기술이 가장 많이 쓰이고 있다. Google Chart API 도 모든 웹 브라우저에서 사용할 수 있고, 많은 수의 시각화 종류를 지원하지만, 한가지 단점이 있다면 외부데이터를 통한 시각화는 지원하지 않는 점이다. 이에 반해 D3.js 는 "HTML5 + 자바스크립트" 를 지원하는 웹 브라우저를 통해 PC, 모바일 등 다양한 매체에서 웹을 통한 시각화 정보를 확인 할 수 있으며, 외부 데이터를 통한 시각화 역시 지원한다 [27]. 마지막으로, D3.js 는 [그림 19] 와 같이 HTML5 기술과 연계된 다양한 그래픽 객체나 차트 생성 기능 등을 제공하며, 객체 변환 및 이동 기능과 함께 다양한 유형과 포맷 의 데이터를 웹 도큐먼트의 엘리먼트(Element)로 바인딩(Binding)하여 웹(Web)에서 적용할 수 있도록 도와준다. 이러한 장점들을 토대로, 본 연구에서는 NV 프로그램을 구현하는데 있어 D3.js 를 적극적으로 사용할 수 있었다.

[그림 19] : D3.js 를 활용한 다양한 종류의 시각화 기법들

### 3.3.2. Program Structure

N-polar Visualization(NV) 의 프로그램 구조는 **[그림 20]** 의 형태를 가지며, 각 부분에 대한 설명은 다음과 같다.

**[그림 20] : N-polar Visualization 프로그램 구조**

Raw Data : 1) UN 투표 데이터, 2)미국 하원 투표 데이터의 가공되지 않은 형태이다.

Processed Data : 3.2. Data Processing 절 에서 설명한 데이터 가공 방법에 따라 데이터 오브젝트(Data Object) 끼리의 1 년을 단위로 유사도를 계산한 CSV 형태의 매트릭스(Matrix) 데이터를 말한다.

layout.html : NV 의 각 View 의 배치를 담당하는 부분이다.

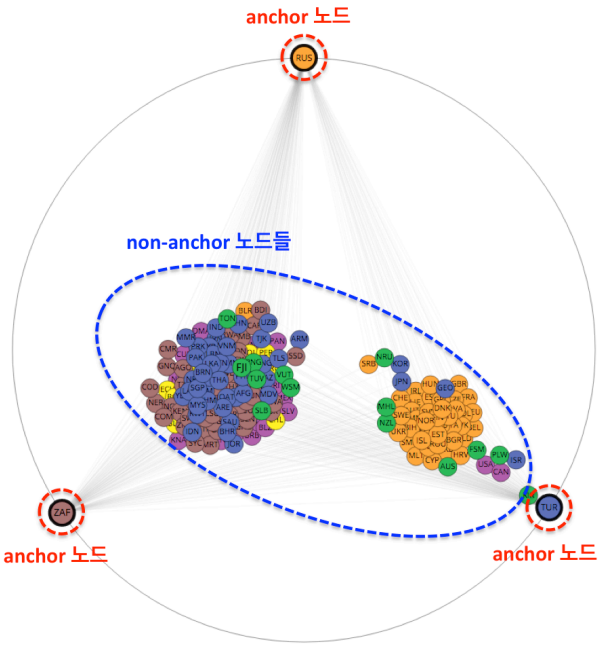
**polar.js** : Processed Data (CSV 파일) 을 자바스크립트로 읽고, D3.js 기술을 이용하여 NV 의 내부 알고리즘 모듈을 시각화하여 보여주는 역할을 수행한다.

**UI.js** : 사용자 인터페이스를 담당하며, Main View, Control View 1, Control View 2, Detail View, Time Slider 5 개의 View 에서 발생하는 이벤트(event)를 담당하는 다양한 함수들이 이에 속한다.

## 3.4. Behaviors of the Nodes and the Links

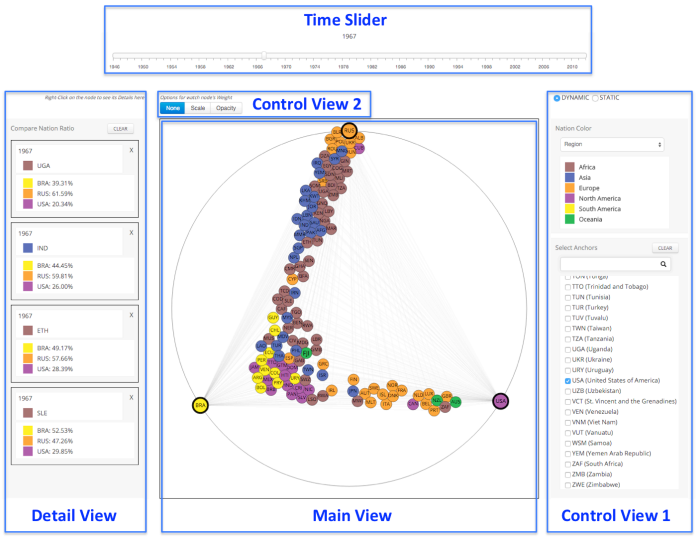
N-polar Visualization(NV) 에서 "anchor" 노드라 함은 빈 원의 둘레의 주변에 고르게 퍼져있는 노드(Data Object)를 의미하며, "non-anchor" 노드는 원 안 쪽에 배치되어 있는 노드들을 말한다. 각각의 non-anchor 노드는 각각의 anchor 노드와 내부적인 정보를 스프링 힘을 기반으로 한, 링크(Link)에 의해서 연결되어 있다. 각 non-anchor 노드 의 위치는 anchor 노드 들이 각각 당기는 힘이 평형에 도달할 때 결정되어진다. **[그림 21]** 의 **Main View** 에서 이를 자세히 확인할 수 있다.

본 연구에서는, anchor 노드 들의 영향력을 향상시키기 위해서, anchor 노드들을 원 둘레에서 움직일 수 있게 하였다. 그로 인해, 사용자는 자기가 흥미롭다고 생각하는 anchor 노드 들을 설정하여 움직여볼 수 있다. 또한, 사용자는 **[그림 22]**, **[그림 23]** 의 **Control View 1** 에서, anchor 노드들을 자유롭게 추가하거나 뺄 수 있다. 이러한 anchor 노드 인터랙션을 통해서, 이 시각화 테크닉은 사용자로 하여금 흥미로운 패턴들의 확인과 중요한 가설들의 검증을 가능하게한다.

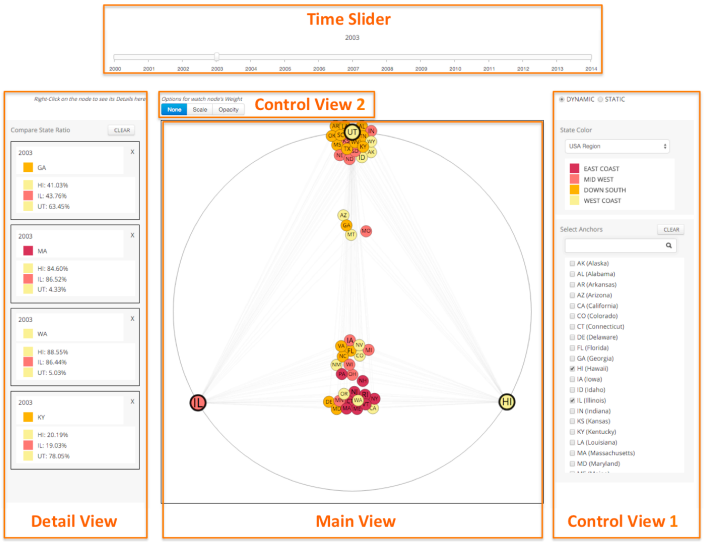



[그림 21] : N-polar Visualization - Main View 에서의 "anchor" 노드 와 "non-anchor" 노드들

NV 에서 각각의 링크들은 스프링과 같이 행동하며, 링크들의 힘은 직접적으로 2 개의 노드 사이의 유사도(Similarity) 값에 비례한다. 그러므로, 노드 2 개의 유사도 값이 크면 클수록, 노드들은 더욱 더 가까워진다. 그로 인해, 노드들이 한 지점에 뭉쳐버리는 현상이 발생하는데, 이러한 현상을 해소하고자 우리는, 적당한 함수의 유사도 값의 스프링 강도를 찾아보기 위해서, 다양한 방정식들을 적용하여 테스트해봤다. 그 중에서, 지수 함수는 선형 함수에 비해 노드들을 더욱더 흩어질 수 있게 하는 경향을 볼 수 있었다. 그래서, 본 연구에서는 지수 함수를 적용하여 사용자들이 NV 시각화를 이용하여, 더욱 더 쉬운 분석을 가능하게 만들었다.




[그림 22] : 1946~2012 년 까지의 UN 투표 데이터를 이용하여 시각화한 NV 의 5 가지 View (Main View, Control View 1, Control View 2, Detail View, Time Slider)




[그림 23] : 2000~2014 년 까지의 미국 하원 투표 데이터를 이용하여 시각화한 NV 의 5 가지 View (Main View, Control View 1, Control View 2, Detail View, Time Slider)

## 3.5. Details-on-demand and Additional Features

본 시각화 연구에서는 노드들간의 관계를 살펴보기 쉽게 하기 위해서 몇 가지 인터랙션 기능을 추가하였다 **[그림 22], [그림 23]**.

첫째, UN 투표 데이터를 사용할 때는 대륙별(i.e. Africa, Asia, Europe, North America, South America and Oceania)로 칼라코드를 주었으며, 미국 하원 투표 데이터를 사용할 때는 미국 지역별(i.e. East Coast, Mid West, Down South, West Coast)로 칼라코드를 설정하였다. 그래서 Control View 1 의 윗 부분을 통해, 자신이 흥미롭게 여기는 대륙과 지역에 대해서 필터링을 해봄으로써, 특정 대륙과 지역에 속한 노드들이 어떤식으로 분포되는지 더욱더 상세한 확인이 가능하다. 이러한 지역구분을 기초로 국가별, 지역별로 색깔을 다르게 표시하면 시각화에서 대륙 혹은 지역간에 어떤 유사성이 있는지를 쉽게 인지할 수 있다.

둘째로 본 시각화는 **[그림 22], [그림 23]** 에서와 같이 Detail View 도 제공한다. Detail View 에서, 사용자는 마우스 오른쪽 버튼을 통해 자신이 흥미로워보이는 non-anchor 노드와 anchor 노드들간에 실제 유사도값을 확인해 볼 수 있다. 이를 통해, 특정 non-anchor 노드가 해당 자리에 배치되어 있는 이유에 대해서 알 수 있다.

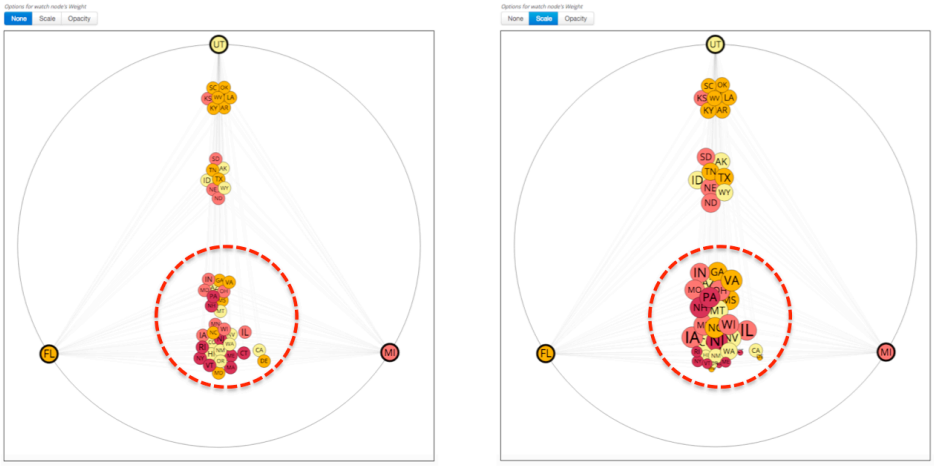
마지막으로, **[그림 22], [그림 23]** 의 Time Slider 를 통해서, 사용자는 시간에 따라 non-anchor 노드들의 레이아웃이 어떻게 바뀌는지를 확인해 볼 수 있다. Time Slider 는 UN 투표 데이터의 경우 1946 년부터 2012 년까지, 미국 하원 투표 데이터의 경우는 2000 년부터 2014 년 까지 1 년 단위로 변화를 확인해 볼 수 있게 설정하였다.

## 3.6. Scale and Opacity

N-polar Visualization 을 사용하다 보면, 비슷한 위치에 놓이는 non-anchor 노드들을 볼 수 있다. 이 때, 사용자는 비슷한 위치에 놓여있는 non-anchor 노드들의 모습을 보고 단순히 기준이 되는 anchor 노드에 대한 실제 유사도 값이 같아, 유사한 위치에 놓이는 거라 생각할 수 있다. 예를들어, non-anchor 노드 A 가 2 개의 anchor 노드에 (0.9, 0.9) 의 유사도 값을 가지고 있고, non-anchor 노드 B 가 동일한 2 개의 anchor 노드에 (0.1, 0.1) 의 유사도를 가지고 있다고 가정해보자. 실제 유사도 값은 확연히 다르지만, 노드 A, B 는 두 anchor 의 상대적인 힘에 의해서 동일한 위치에 놓일 것이다. 그래서 비슷하거나 동일한 위치에 놓여있는 non-anchor 노드들에 대한 구분을 명확히 하기 위해, Scale(크기) 과 Opacity(불투명도) 기능을 추가하여 이를 해결하고자 하였다.

화면에 표시되는 non-anchor 노드의 Scale 과 Opacity 는 Weight 값의 [최소, 최대] 값들을 정의역(Domain)으로 설정 후, 내삽(Interpolation) 하여 각 Weight 값을 [0, 1] 사이의 값들로 변경하였다. Scale 기능이 활성화된 상태에서, non-anchor 노드의 크기는 anchor 노드 들간의 유사도 합의 평균값으로 설정되어있어 유사도의 평균값이 클 수록, non-anchor 노드의 지름은 커지며, 유사도 평균값이 작을 수록, non-anchor 노드의 지름은 작아진다. Scale 기능과 비슷한 맥락으로, Opacity 기능은 non-anchor 노드의 앵커노드들 간의 유사도의 평균값이 클수록, 노드가 선명하게(불투명하게) 나타나지만, 유사도의 평균값이 작다면 투명하게 나타난다.

해당 값을 사용하여 노드의 Scale 로 변화(Variation)를 표현할 경우 { value * radiusScaleFactor + 5.0 } 으로 계산했으며, 노드의 Opacity 로 표현할 경우 { value * (1.0 - opacityScaleFactor) + opacityScaleFactor} 의 방식으로 결정하였다.

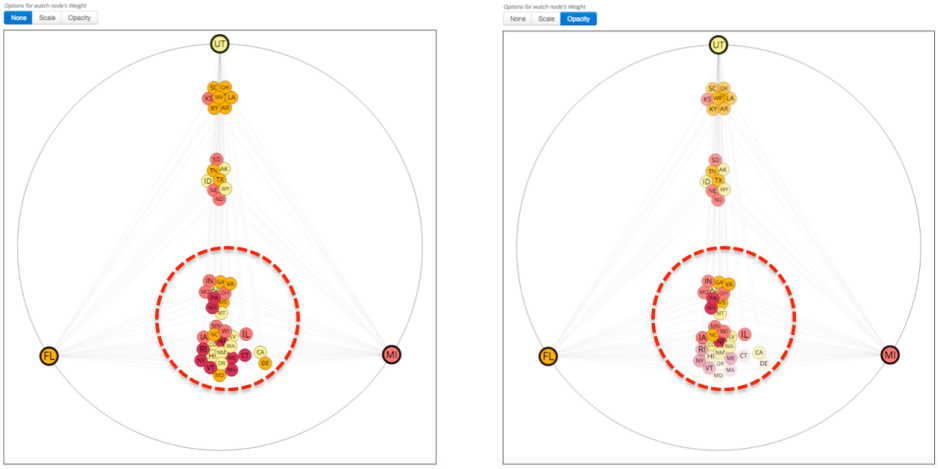



(a) None (b) Scale

[그림 24] : Scale 기능을 통한 non-anchor 노드들에 대한 구분

[그림 24], [그림 25] 는 None, Scale, Opacity 기능이 활성화된 상태를 보여준다. 사용자는 [그림 22], [그림 23] Control View 2 의 None, Scale, Opacity 버튼을 클릭하여 non-anchor 노드에 변화를 줄 수 있다.

[그림 24] 의 (a) 와 같이 None 이 활성화 되어 있을 때는, 빨간 점선표시 안과 같이 클러스터가 생길 때 각 노드의 유사도 값이 어떻게 구성되어 있는지 시각적으로 알기 힘든 어려움이 있다. 하지만, [그림 24] 의 (b) 와 같이 Scale 기능을 활성화시키면, 각 노드들의 크기에 따라 실제 유사도 값이 어떤식으로 구성되어 있는지 사용자로 하여금 빠른 인식을 가능하게 해준다. [그림 24] 의 (b) 를 자세히 보면 노드의 지름이 큰 집단, 중간 정도의 집단, 아주 작은 집단 3 개로 구성됨을 볼 수 있다. 노드의 지름이 크다는것은 다른 non-anchor 노드들보다 3 개의 anchor 노드에 대해서 높은 유사도 값을 가지는 것을 의미하며, 노드의 지름이 작다는건 anchor 노드 들간에 낮은 유사도 값을 가진다고 할 수 있다.




(a) None (b) Opacity

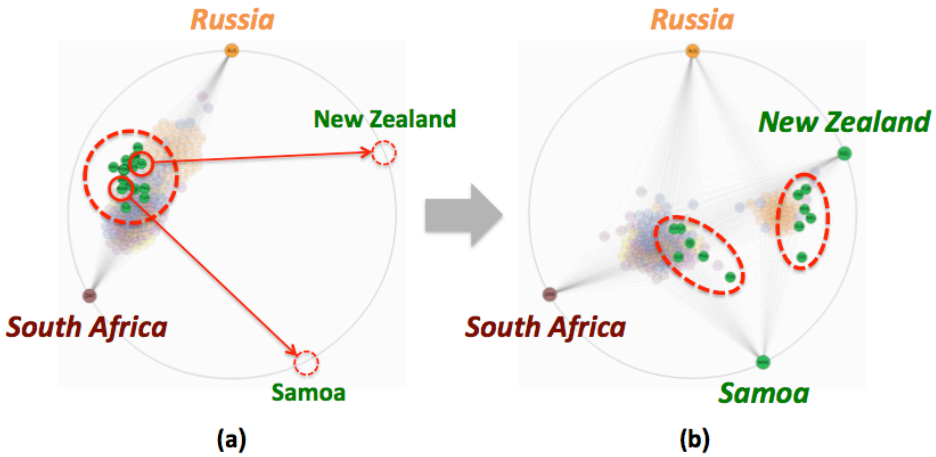
[그림 25] : Opacity 기능을 통한 non-anchor 노드들에 대한 구분

[그림 25] 의 (b) 의 경우는 Opacity 를 통해서 non-anchor 노드들에 변화를 준 것이며, Scale 기능과 비슷한 맥락을 따른다. 각 노드들의 투명도에 따라 non-anchor 노드의 실제 유사도 값이 어떤식으로 구성되어 있는지 사용자로 하여금 빠른 인식을 가능하게 해준다. [그림 25] 의 (b) 를 자세히 보면 선명한 노드와, 적당히 투명한 노드, 투명한 정도가 높은 노드 3 개의 집단으로 구성됨을 볼 수 있다. 노드의 투명도가 낮다는것(노드의 선명도가 높다는것) 은 3 개의 anchor 노드 들간에 타 non-anchor 노드 들보다 높은 유사도 값을 가지는 것을 의미하며, 노드가 투명정도가 높다는건 anchor 노드 들간에 낮은 유사도를 가진다고 할 수 있다.

두가지 기능(Scale, Opacity)을 통하여 사용자는 같은 위치에 놓여있는 non-anchor 노드가 anchor 노드 들간에 높은 유사도로 해당 위치에 있는지, 낮은 유사도로 해당 위치에 있는지 시각적으로 확인이 가능하다.

## 3.7. Interactive Exploration

본 연구에서 non-anchor 노드 들은 각각의 anchor 노드들에 대하여 비슷한 유사도 를 가질 경우, 특정 지점에서 뭉치려는 경향을 띈다. 이 때, 뭉쳐있는 상태를 클러스터형 레이아웃이라 명명할 수 있는데, 이 때, 사용자는 클러스터 안에 있는 특정 노드들 몇 개를 선택해 anchor 노드로 둠으로써, non-anchor 노드들이 이룬 클러스터형 레이아웃이 어떻게 변하는지 볼 수 있다. 예를 들어, **[그림 26]** 을 보면 Russia 와 South Africa 를 anchor 노드로 설정했을 때, 대부분의 UN member 들이 한 지점에 뭉쳐져 있는것을 볼 수 있다. 이 뭉친 지점에서, 특정 오세아니아 대륙(녹색 노드)의 국가들을 하이라이팅 해서 보아도, **[그림 26]** 의 (a) 와 같이 한 지점에 뭉쳐진 클러스터형 레이아웃을 띈다. 여기서, Oceania 대륙의 노드들 중 2 개(New Zealand, Samoa)를 선택해서 또 다른 anchor 로 추가 함으로써, 기존에 1 개의 클러스터로 뭉쳐져 있었던 Oceania 대륙의 노드들이, **[그림 26]** 의 (b) 와 같이 2 개의 클러스터로 나뉘어짐을 확인할 수 있다.




**[그림 26] : Oceania 대륙 1 개의 클러스터가 2 가의 anchor 노드를 추가로 설정함으로써 2 개의 클러스터로 나뉘어지는 사례**

# 제 4 장 Pattern Study and Case Study

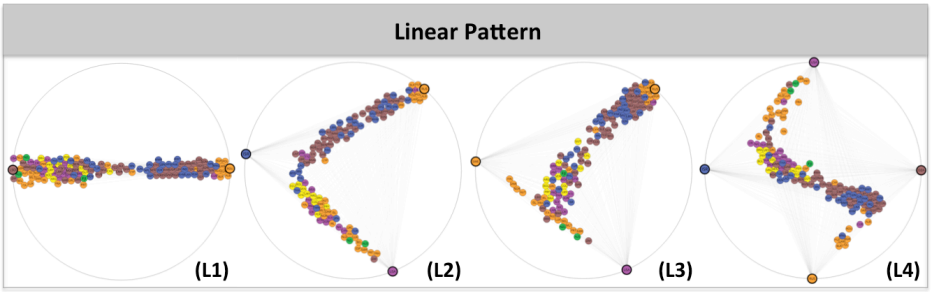
본 장에서는 N-polar Visualization 의 유용성(Usefulness)과 효율성(Effectiveness)을 평가해보기 위해서 UN 투표 데이터와 미국 하원 투표 데이터를 통해 일련의 조사를 수행하였다. 첫째, 빈번하게 보이는 패턴들을 찾고 이를 정의할 수 있었고 둘째, 두 가지 사례 연구를 통하여 시간의 흐름에 따른 특정한 레이아웃이 실제 역사적 사건과 일치하는 결과를 볼 수 있었다.

## 4.1. Pattern Study

해당 시각화 테크닉이 얼마나 유용하게 사용될 수 있는지를 규명하기 위해서, UN 데이터셋을 이용하여, 어떤 종류의 패턴이 나타나는지 확인해보았다. ("패턴" 이라 함은 선택되어 있는 특정한 anchor 노드들을 기준으로 non-anchor 노드들의 분포 형태를 뜻한다.) 이러한 조사를 통해서, 많은 수의 패턴들을 찾아내었고, 이를 2 개의 그룹(Linear Pattern, Clustering Pattern)으로 분류하였다.

### 4.1.1. Linear Pattern

Linear Pattern (선형 패턴) 은 소위 선 모양의 패턴을 말하며, non-anchor 노드들이 특정한 지점에 집중되어 있지 않는 형태이다. 이는, 유사도(Similarity)가 낮은 두개의 anchor 노드를 양 끝으로 하고, non-anchor 노드들이 일렬로 고르게 분포되어 있는 경우에 주로 나타난다. 그 중에서, 특히 (L1) 모양의 패턴을 종종 볼 수 있었으며, (L2), (L3) 의 경우는 (L1) 패턴에서 또 다른 한개의 non-anchor 노드를 3 번째 anchor 노드로 두었을 때 나타난다. 마지막으로 (L4) 의 경우 (L2) 에서 다른 한개의 non-anchor 노드를 4 번째 anchor 노드로 설정하면 발견되는 패턴이다 **[그림 27]**. 하지만, **[그림 27]** 의 (L4) 의 경우 사용자가 패턴을

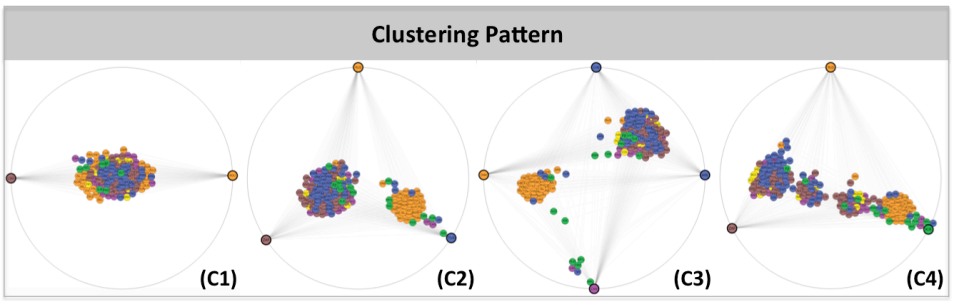



[그림 27] : Linear Pattern (L1, L2, L3, L4)

해석하는데 있어, 어려움이 발생할 수 있다. 왜냐하면, anchor 를 어떤 방향에 놓느냐에 따라 (L4) 가 기존과는 다른 패턴으로 보일 수 도 있기 때문이다.

## 4.1.2. Clustering Pattern

Clustering Pattern (군집 패턴) 은 non-anchor 노드들이 특정 지점에 원형으로 뭉쳐져 있는 형태이며, 다수의 non-anchor 노드들이 anchor 노드들 각각에 대해서 유사도가 비슷함과 동시에 높은 유사도 값을 가지면 (C1) 패턴을 볼 수 있다. 때때로 (C2), (C3), (C4) 와 같은 2~4 개의 군집이 보이는 Clustering Pattern 도 발견할 수 있었다 [그림 28].




[그림 28] : Clustering Pattern (C1, C2, C3, C4)

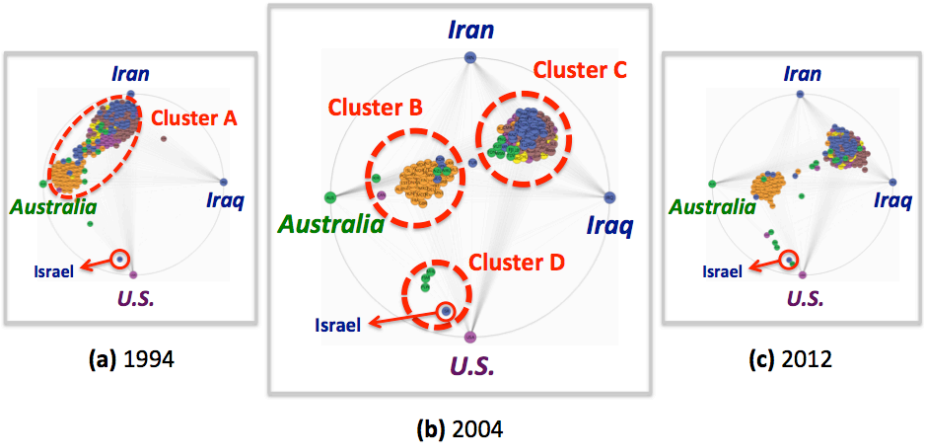
## 4.2. Case Study

N-polar Visualization 의 효율성을 살펴보기 위해서, 2 개의 데이터셋(1.UN 투표 데이터, 2.미국 하원 투표 데이터) 을 가지고 세계에서 일어났던 실제 역사적 사건에 대한 2 개의 사례 연구를 진행하였다.

### 4.2.1. 이라크 전쟁 (2003 ~ 2011)

본 연구에서는 UN 투표 데이터를 이용한 첫 번째 사례 연구로 2000 년 대에 세계적으로 큰 사건이라 할 수 있는 Iraq War(이라크 전쟁)에 대한 스터디를 진행하였다. 그리고, Iraq War (2003~2011) **(a) 전-1994, (b) 중-2004, (c) 후-2012** 3 가지의 시간대 별로 구간을 나누고, 각각의 레이아웃에 실제 역사적 사건을 대입해 봄으로써 예측할 수 없었던 놀라운 사실들을 알아 볼 수 있었다. 사례연구 에서는, Iraq, Iran, Australia, USA 4 개의 국가를 anchor 노드로 배치하였다. Iraq War (2003~2011) 중에 Iraq, Iran 는 Iraqi Force 였으며, Australia, USA 는 U.S. 연합 이었다. anchor 노드 4 개를 동일한 거리로 **[그림 29]** 의 (a), (b), (c) 처럼 배치시켰다. 그런 다음, non-anchor 노드들이 어떤 패턴의 레이아웃을 가지는지 살펴보았다. 여기서, (a), (b), (c) 각각의 그림에 대한 흥미로운 패턴들을 볼 수 있었다.

Iran-Iraq War (1980~1988) 가 끝난 후의 시점인 1994 년 **[그림 29]** 의 (a) 를 보면 Iran 과 Australia anchor 노드에 걸친 큰 Cluster A 가 보이는 것을 알 수 있다. 이는, 거의 대부분의 국가들이 USA 와 Iraq 가 아닌, Iran 과 Australia 에 더 가까운 것을 볼 수 있었다. 이는, Iran-Iraq War 관련 신문기사 [28] 에서, “대부분의 국가들이 Iraq 가 아닌 Iran, Australia 에 더 우호적인 관계를 가지는 것” 으로써 검증할 수 있었다.




[그림 29] : 역사적 사건인 Iraq War (2003~2011) 전, 중, 후 에 대해서 UN member 들 간의 관계를 밝혀낸 사례 연구

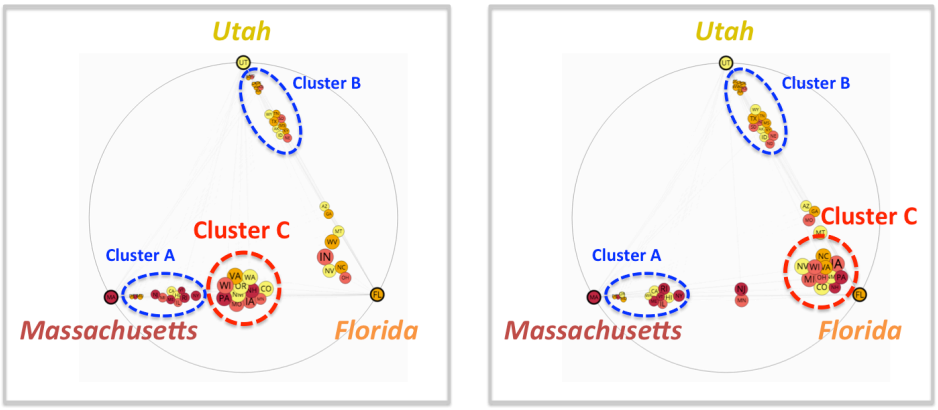
[그림 29] 의 (b) 는 Iraq War (2003~2011) [26] 도중의 2004 년에 대한 anchor 노드들과 UN Member 들간의 관계에 대해서 보여주고 있다. [그림 29] (a) 에서 보였던 Cluster A 가 (b) 에서는 3 개의 Cluster B, C, D 로 나뉘어지는 것을 확인할 수 있었다. Cluster B 는 European 국가들로 구성되어 있는 반면에, Cluster C 는 African, Asian, North American, South American 대륙의 국가들이 주를 이루었다. 이러한 패턴들은 UN Member 들의 투표 패턴이 대륙별로 비슷하게 움직이고 있다는 것을 의미한다. 또한 Cluster B 는 Australia 와 USA 와 가까운 반면에 Cluster C 는 Iran 과 Iraq anchor 노드와 가까운 경향을 보였다. European 대륙의 국가들은 U.S. 연합 에 높은 투표 유사도를 보이는데 반해, African, Asian, North American, South American 대륙의 국가들은 Iraqi Force 국가인 Iran, Iraq 에 높은 투표 유사도를 보였다. 또 한가지 흥미로운 패턴이 있다면, Iraq War 중인 2004 년에 U.S. 연합을 포함한 다수의 UN 국가들이 상대적으로 USA 와 낮은 투표 유사도를 가진다는 점이었다. 또한, U.S. 연합국가들에 대해서도 USA 가 아닌, Australia 쪽으로 Pro-U.S. 국가들의 투표 유사도가 더 높은 경향을 보였다. 마지막으로, [그림 29] (b) 에서

Cluster B, C 에 속하지 않는 Cluster D 를 찾을 수 있었는데, 여기에는 Israel 과 Palau 를 포함한 4 개의 국가가 USA 와 높은 투표 유사도를 보인다는 것을 알 수 있었다. 실제로 Iraq War 시기를 언급하고 있는 신문기사 [29] 에서, "Iran 과 Iraq 는 동맹국이었고, 유럽 대륙을 제외한 대부분의 국가들이 U.S. 연합이 아닌 Iran 과 Iraq 에 우호적인 성향" 을 보이는 것을 확인할 수 있었다.

Iraq War(2003~2011)가 끝난 뒤, **[그림 29]**의 (c) 2012 를 보면, 2004 때 와 비슷한 양상의 패턴을 볼 수 있다. 또한 이스라엘 국가는 2012 년 에도 역시 여전히 미국과의 투표 결과가 계속 유사하다. 흥미로운 사실은 3 개의 시간 때를 순차적으로 보아도, 미국과의 투표 결과가 꾸준히 유사하다는 점이다. 이는, 신문기사 [30] 에서 언급한, "이스라엘과 미국은 1990 년대 이후 우호관계를 지속하고 있다는 점"과 일맥상통하고 있음을 알 수 있었다.

### 4.2.2. 미국 대통령 대선 (56 회, 57 회)

두 번째 사례연구는 미국의 하원 투표 데이터를 이용하여, 미국의 대통령 대선이 있었던 (a) 2008, (b) 2012 두 가지의 시간대 별로 구간을 나누고, 각각의 레이아웃에 실제 역사적 사건을 대입해 봄으로써 흥미로운 사실들을 알아볼 수 있었다 **[그림 30]**. 하원 투표 데이터를 활용한 사례 연구에서는, Florida, Utah, Massachusetts 3 개의 State 를 anchor 노드로 배치하였다. 미국의 대통령 선거에서 최근 10 년간 Florida 는 Swing State (경합주) 이고 Utah 는 Republican State (공화당) , Massachusetts 는 Democratic State (민주당) 투표를 많이 받은 곳이었다 [31] [32]. 위에서 언급한 3 개의 anchor 노드를 동일한 거리로 **[그림 30]** 의 (a), (b) 처럼 배치시켰다. 그런 다음, non-anchor 노드들이 어떤 패턴의 레이아웃을 가지는지 살펴봄으로써 (a), (b) 각각의 그림에 대한 흥미로운 패턴들을 확인해 볼 수 있었다.




(a) 2008 (b) 2012

[그림 30] : 미국의 56 회(2008 년), 57 회(2012 년) 대통령 대선에 대해서 미국의 50 개 State 들 간의 관계를 밝혀낸 사례 연구

일단 전반적으로 State 의 지리적 위치에 상관없이 다양한 지역의 State 들이 클러스터로 구성됨을 확인할 수 있었다. 이러한 패턴은 각 State 들의 미국 하원의 투표 패턴이 지리적 위치에 상관없이 움직이고 있다는 것을 의미한다.

미국 56 회 대통령 선거 시점인 2008 년 [그림 30] 의 (a) 를 보면 Cluster A 에는 Delaware, Connecticut, Maryland, Maine, New York, Hawaii 등의 State 들이 있으며, 이들은 Massachusetts anchor 노드에 가까이 분포하는 경향을 보인다 [33]. Cluster A 에 속하는 non-anchor 노드 들은 2008 년 대선 때 민주당(Democratic) 표가 많았던 곳으로, 해당시기에 민주당 표가 많은 곳이라 할 수 있는 Massachusetts 와 높은 투표 유사도를 보인다. 비슷한 맥락으로, (a) 의 Cluster B 에는 South Carolina, North Dakota, Texas, Indiana 등의 State 들이 있으며, 이 들은 실제 2008 년 대선 때 공화당(Republican) 표가 많이 집계되었으며, Utah anchor 노드와 상대적으로 높은 투표 유사도를 가지는 것을 볼 수 있다 [33].

[그림 30] 의 (a) 2008 대통령 대선 이후로, (b) 2012 를 보면, Cluster A 와 Cluster B 는 2012 년에 와서도 비슷한 패턴을 가짐을 볼 수 있다. 이는 "Cluster A, Cluster B 에 속하는 State 들이, 2008 년 대통령 대선 때 Democratic, Republican 의 투표성향이 강했던 State 들은 2012 년에 와서도 그 성향이 변하지 않는" 기사 내용 [34] 과 일치하는 것을 확인해 볼 수 있었다. 또한, 2008 년의 Cluster A 에 속해있던 New Jersey, Michigan 과 Cluster B 에 속해있던 Arkansas 를 제외하고는 2008 년에서 2012 년으로 시간이 흘러도 거의 비슷한 양상을 가지는 것 역시 확인하였다.

마지막으로 [그림 30] (a) 2008 년에 Florida 와 Massachusetts anchor 노드 사이의 Cluster C 를 볼 수 있다. Cluster C 에는 Wisconsin, Pennsylvania, New Mexico, New Hampshire 등을 포함한 11 개의 State 가 있으며 Massachusetts 와 Florida anchor 노드와 가까운 관계에 있고, Utah 노드 에는 먼 관계에 있음을 알 수 있다 [35]. 미국 56 회 대통령 선거(2008)에서 57 회 대통령 선거가 일어난 [그림 30] (b) 2012 년으로 갔을 때 Cluster C non-anchor 노드의 구성은 거의 바뀌지 않는다. 하지만 여기서 흥미로운 사실은, 클러스터의 위치면에서 Florida 쪽으로 옮겨짐에 따라 2012 년에는 Cluster C 에 속해있는 State 들이 Florida 와 큰 유사 관계가 나타난다는 것이었다. Cluster C 에 속하는 Wisconsin, Pennsylvania, New Mexico, New Hampshire 는, 2008 년 미국 대통령 대선때는 Democratic State 인 경향을 보였지만, 2012 년에는 2008 년 대비 더 이상 Democratic State 라고 단정지을 수 없는 투표 결과를 확인할 수 있었다. 실제로 신문기사 [36] 에서는 "Wisconsin(-3.7%), Pennsylvania(-2.7%), New Mexico(-2.4%), New Hampshire(-2.1%) 결과와 같이, 2008 년에는 Democratic 였던 State 들은 2012 년에는 (-) 방향으로 투표결과" 를 볼 수 있었다. 이는 Cluster C 에 속한 대부분의 State 들이 2008 년 대비 2012 년에 Democratic State 들과 상대적으로 낮은 유사관계를 보인다고 말할 수 있다. 마지막으로 Cluster C 는 2008 년, 2012 년 모두 Scale 의 노드 지름이 다른 State 들에 비해 크게 나타나므로, 다른 non-anchor

노드에 비해 3 개의 anchor 노드(Florida, Utah, Massachusetts) 와 평균적으로 높은 유사도를 가지는 것 역시 확인해볼 수 있다.

### 4.2.3. 정리 및 한계

이와 같은 일련의 두 가지 사례 연구를 통해서 N-polar Visualization(NV) 테크닉이, 시간의 흐름에 따른 데이터에서 숨겨진 인사이트를 찾는데 중요한 역할을 한다고 볼 수 있었다. 또한, 사례 연구들을 통해서 발견한 인사이트와, 기존에 데이터만으로는 알기 힘들었던 패턴들이 실제 발간 되었던 신문기사 (ex: The New York Times, The Washington Post, The Guardian 등)의 내용과 일치하는 결과 역시 확인함으로써 NV 의 효율성을 검증해 볼 수 있었다.

하지만, 본 사례 연구에서는 두 가지의 잘 된 경우에 대해서만 서술한 면이 있다. 전쟁 관련 역사적 사건들을 NV 로 시각화 한다고 가정한다면, 전체 의제와 법안을 볼 것이 아니라, 전쟁과 관련되어 있는 의제와 법안만을 선별하여 시각화 했어야 했다. 왜냐하면, UN 투표 데이터의 의제들은 하나의 주제가 아닌, 다양한 주제를 가지고 있고, 미국 하원 투표 데이터 역시 다양한 법안들로 구성되어 있기 때문이다. 그래서, 가공한 2 개의 데이터셋(UN 투표데이터, 미국 하원 투표 데이터) 으로 다른 역사적 사건들을 NV 테크닉에 적용한다면, 실제 역사적 사실과 일치하지 않는 경우도 존재할 것이라 사료된다. 결론적으로, 본 연구의 사례 연구에서 진행한 전체 데이터가 아니라, 실제 역사적 사건과 관련성이 높은 의제와 법안만을 선별하여 진행 했다면 더 타당(Validation)한 연구가 되었을거라 본다.

# 제 5 장 결론 및 연구의 의의

## 5.1. 연구 요약

본 연구는 기존의 선행연구들이 가지는 한계점을 개선하는 새로운 시각화 테크닉을 제안한 연구이다. 연구 초반에는 Data Object 각각의 관계를 효과적으로 시각화하려는 선행 연구들을 정리하고, 그들이 가지는 한계점을 설명하였다. 그리고 이러한 한계점들을 RadViz (Radial Visualization) 시각화 테크닉의 메커니즘(Machanism) 을 활용하여 개선점을 찾았으며, 동시에 다양한 이점들도 발견할 수 있었다.

제안한 테크닉의 이름은 다극 시각화(N-polar Visualization)라 명명했으며, 내부 시각화 알고리즘과 사람들이 사용할 수 있을 정도의 프로토타입을 구현하고, 추가적으로 필요한 인터랙션 기능과 사용자 인터페이스(UI)를 개발하였다. 그리고 구현한 시각화 테크닉의 실행가능성(Feasibility)를 평가해보기 위해서, Data Object 간에 높은 관련성이 있는 특정 데이터를 적용한 일련의 예비 조사를 수행하고, 제안한 테크닉이 특정한 데이터에 관계없이 일반적(Generalization)으로 사용될 수 있는지, 추가적인 데이터셋을 가공하여 적용할 수 있었다.

본 연구에서는 최종적으로 제안한 N-polar Visualization 이 얼마나 유용하고 효율적으로 사용될 수 있는지를 규명하기 위해, 패턴연구(Pattern Study)와 사례연구(Case Study)를 진행하였다. 첫째, UN 투표 데이터를 활용하여 어떤 종류의 패턴이 주로 발견되는지 확인하는 과정을 가졌다. 그래서 다양한 패턴들을 발견할 수 있었고, 이를 2개의 그룹(Linear Pattern, Clustering Pattern)으로 분류할 수 있었다. 둘째, 실제 일어났던 역사적 사건들을 토대로 사례연구를 진행하여 시간의 흐름에 따른 데이터에서 숨겨진 인사이트를 찾는데 있어 제안한 테크닉의 중요성을 확인해 볼 수

있었다. 또한, 발견한 인사이트와 기존에 데이터만으로는 알기 힘들었던 패턴들이 실제 발간되었던 신문기사와 일치하는 것을 확인할 수 있었다.

본 연구를 통해, 사용자는 데이터 오브젝트 각각에 대해서 연관성이 높은 데이터를 더 쉽게 이해할 수 있으며 기존의 선행연구로는 알기 힘들었던 새로운 인사이트 역시 얻을 수 있을 것으로 기대한다.

## 5.2. 연구의 시사점

우리는 단순히 정보화 시대가 아닌 정보과부화(Information Overload) 시대라 할 수 있는 빅 데이터(Big Data)의 시대에 살고 있다. 정보의 양이 워낙 압도적이다 보니, 적절한 방법을 사용해서 숨은 의미를 이끌어 내지 않으면 그 정보는 데이터로서 전혀 유용하지 않게 된다. 다행히 인간은, 오감 중 시각이 가장 발달한 생물이다. 어떤 정보를 숫자만 길게 나열한 표로 만들면 일부 사람들만이 그 안에 있는 규칙을 찾아내지만, 막대 차트 같은 시각적 표현으로 바꾸면 아이들도 쉽게 규칙을 알아차릴 수 있다. 이처럼 시각화는 표현력이 강력해서 다른 사람에게 단 시간에 의견을 전달하기에 가장 좋은 수단이라고 할 수 있다. 또한, 정보에서 미처 알지 못했던 규칙과 동향을 찾아내고, 새로운 방법을 통해서 새로운 인사이트를 볼 수 있도록 도와준다. 다시말해서, 수 많은 정보들을 시각적으로 묘사하고 필요한 정보를 효율적이고 명확하게 제공하는 정보시각화(Information Visualization)의 필요성이 빅 데이터 시대에 있어 아주 중요하다는 의미이다.

본 연구에서는 이러한 시대적 흐름에 발맞추어 정보시각화 학술분야에 존재하지 않는 새로운 시각화 테크닉인 다극 시각화(N-polar Visualization)를 제안하였다. N-polar Visualization(NV)은 데이터 오브젝트(Data Object) 의 관계를 시각화한 다른 테크닉들과 비교해서 많은 장점을 가지고 있다. 무엇보다도, 다중의 인터랙티브한 anchor 노드가 가장

큰 특징이라고 할 수 있다. 이는 사용자들로 하여금 자신이 흥미롭다고 생각하는 데이터 오브젝트 몇 개를 anchor 노드를 선택함으로써, 복잡한 데이터 오브젝트들 간의 관계를 풀어낼 수 있다는 점이다. 또한, 사용자의 anchor 노드에 대한 조작을 유연하게 하기위해서, anchor 노드를 원의 둘레에서 자유롭게 움직일 수 있게 하였다. 이를 통해, 사용자는 관심이 있을만한 데이터 오브젝트 들을 anchor 노드로 설정하여, 원 안쪽의 non-anchor 노드들이 선택된 anchor 노드 들과 어떤 관련성을 가지는지 살펴볼 수 있으며, 움직이는 anchor 노드에 따라 달라지는 non-anchor 노드의 패턴 변화 양상 역시 확인해 볼 수 있다.

본 연구를 통해서, 사용자는 데이터 오브젝트 각각에 대해서 관계가 복잡한 데이터를 가지고 있을 때, NV 테크닉을 이용하여, 가진 데이터를 더 쉽게 이해할 수 있으며, 기존의 시각화 방법으로는 알기 힘들었던 새로운 인사이트 역시 얻을 수 있을 것으로 기대된다. 예를 들어, A 라는 사용자가 학과 50명의 친구들 간에 전공성향에 대한 유사도를 계산한 매트릭스를 만들었다고 가정해보자. A가, NV 를 이용하여 자기 자신과 친구 B 를 anchor 노드로 설정한다면 학과 안의 나머지 48명의 친구들이 A 와 B 누구에게 전공이 상대적으로 더 유사한지 확인해 볼 수 있다. 또한, A 와 B 말고도 C 와 D 친구를 anchor 노드로 배치하여, A, B, C, D 라는 4명의 기준이 되었을 때는 상대적으로 나머지 46명의 친구들은 상대적으로 누구에게 가깝고, 누구에게 먼지에 대한 변화 양상을 확인할 수 있다. 이는 기존 데이터 오브젝트간의 관계를 시각화한 선행 연구들과는 큰 차별점으로 볼 수 있으며, 사람과 사람간의 복잡한 관계를 풀어내는것 뿐만아니라, 부서와 부서, 기업과 기업등의 데이터 오브젝트끼리 상호 관계를 가지고 있는 경우에 대해 매우 유용하게 사용 될 수 있다. 또한, 최근 소셜 미디어의 발전으로 개인이나 조직간의 관계에 대한 관심이 많아지고 있는 추세로 볼 때, NV의 활용가능성은 더 크다고 볼 수 있다.

기술적 측면에서, NV는 HTML5 + 자바스크립트 기반의 D3.js 기술을 사용하여 만든 데이터 시각화 프로그램이다. 이는 웹(Web) 기반의 응용 프로그램이므로, 웹 브라우저만 있으면 초보자들도 손쉽게 몇 번의 마우스

클릭만으로 데이터 분석을 할 수 있는게 특징이다. 또한, D3.js는 데이터 분석의 중요성을 알지만 비용 등의 문제로 상용 패키지의 도입이 부담스러운 소규모 기업이나 개인 등이 이를 활용하는 방법을 통해, 비교적 큰 비용 없이 데이터 분석 환경을 구축할 수 있는 장점이 있다. 왜냐하면 D3.js는 무료 라이센스로 배포되어, 추가 비용 없이 영리 목적이든 비 영리 목적이든 코드를 마음대로 사용, 수정, 보강할 수 있기 때문이다. 그리고, NV가 웹 표준기술로 만들어졌다는 것은 이 시각화를 누구나 운영체제(윈도우, 맥, 리눅스)와 장비의 종류(노트북, PC, 스마트폰, 태블릿)에 관계없이 열람하고 경험할 수 있으며, 로컬 기반에 데이터 분석 시스템을 보다 사용자 접근성이 용이한 웹 기반의 데이터 분석 시스템으로 개선할 수 있다는 산업적 의의가 있다.

또한, 본 연구에서는 NV 시각화를 구현하는데 있어 공공데이터를 사용하였다. 공공데이터는 정부 또는 공공기관이 보유하고 있는 데이터를 의미한다. 2009년 미국의 오바마 대통령은 data.gov 웹사이트를 구축하고, 정부데이터를 일반 사람들에게 공개하였으며, 영국 정부는 공공데이터를 링크드 데이터 방식으로 제공하였다. 이와 같이, 여러 국가에서 국민이 공공데이터를 활용할 수 있도록 공공데이터의 개방 및 인프라를 구축하는데 많은 노력을 기울이고 있으며, 정부에서도 공공데이터 활용이 대국민 서비스와 직결되어 있음을 인지하고, 공공데이터 활용을 정부 정책으로 적극 추진하고 있다. 본 연구에서는 공공데이터의 중요성을 알리기 위해, 일반 국민이 관심을 가질 만한 공공데이터로 UN 투표 데이터와, 미국 하원 투표 데이터 2개를 사용하였다. 이 2개의 데이터는 외교와 정치, 교육 목적으로 활용 가치가 클 뿐 아니라 데이터를 쉽게 얻을 수 있는 장점이 있다. 이와 같이, 본 연구에서는 어느누구나 공공데이터를 얻고 활용할 수 있다는 것을 보여준 점 역시 산업적으로 의의가 있다.

마지막으로, 데이터 오브젝트(개인 or 조직)들 간에 관계(Relationship)를 이해하는 것은, 최근들어 정보시각화 분야에서 주목하고 있는 중요한 과제이다. 이처럼, 주목하는 분야에서 기존의 시각화 테크닉들이 가지는 한계점들을 개선하고, 이를 다양한 데이터셋을 통해

일반화(Generalization)함과 동시에, 일련의 패턴 연구와 사례 연구를 통해서 유용성(Usefulness)과 효율성(Effectiveness)을 검증한 측면은 정보시각화 분야에서의 학술적 의의가 있다.

## 5.3. 연구의 한계 및 제언

본 연구에서는 기존에 존재하지 않는 새로운 시각화 테크닉인 N-polar Visualization(NV)을 제안하고, 이에 대한 유용성과 효율성을 밝혀내었지만, 한편으로는 몇 가지 한계점 또한 존재하였다.

첫째로, NV 테크닉은 사용자 스스로가 anchor 노드(기준 Data Object)를 설정해야만 한다는 것이다. 이는 시각화 하려는 데이터에 사전지식이 없거나 부족한 사람들에게 어려운 요소로 다가올 수 있다. 물론, anchor 노드를 사용자가 지정할 수 있다는 요소는 사람들에게 자유도를 높여줌으로써, 스스로가 분석가가 되어 흥미로운 패턴과 인사이트를 찾아나갈 수 있다는 의미로 볼 수도 있다. 하지만, 데이터에 사전지식이 없거나 부족한 사용자에게도 anchor 노드설정을 위임해야만 한다는 점은 본 연구의 한계라 볼 수 있다.

둘째, anchor 노드의 갯수가 증가하면 할수록 non-anchor 노드들이 가운데로 뭉치는 현상이 발생한다는 점이다. 본 연구에서는, N-polar Visualization 의 내부 스프링 강도를 다양한 방정식들을 적용하여 노드들을 최대한 흩어질 수 있는 작업을 수행하였다. 그 중에서, 지수함수는 선형 함수에 비해 노드들을 더욱 더 흩어질 수 있게 하는 경향을 볼 수 있었다. 하지만 이 역시, 설정된 anchor 노드가 많아지면 non-anchor 노드들의 뭉쳐버리는 현상을 완전히 해결할 수는 없었다. 향후에는, 유사도 값의 내부 스프링 강도에 대한 정교한 연구를 통해, anchor 노드의 갯수가 증가할 수록 가운데로 뭉치는 non-anchor 노드들을 최대한 흩어질 수 있는 방법을 마련해야 한다고 생각한다.

마지막으로, anchor 노드를 원 둘레의 어느 방향에 놓느냐에 따라 패턴이 급격하게 변화하므로, 사용자가 일련의 태스크를 수행하는 와중에 중요한 인사이트를 놓칠 수 있는 점이다. 예를 들어, 4 개의 anchor 노드 A, B, C, D 를 원 둘레에 차례대로 12 시, 3 시, 6 시, 9 시 의 방향으로 설정한다고 가정해보자. 이 때, non-anchor 노드의 패턴은 'S' 형태를 띄는 것을 볼 수 있다. 하지만, anchor 노드의 위치를 D, B, C, A 순서로 바꿔 놓거나, anchor 노드를 동일하지 않은 간격으로 배치한다면 'S' 자가 아닌 전혀 다른 패턴이 만들어질 수 도 있다. 이처럼, 사용자는 anchor 노드의 위치를 일일히 바꿔보면서 확인을 해야하는 어려움이 있다. 따라서, 추후에는 사용자가 인사이트를 찾는데 있어 쉬운 이해를 돕도록 하는 적절한 anchor 노드의 배치에 대한 논의가 이루어져야 한다고 생각한다.

# 참고 문헌

[1] Card, Stuart K., Mackinly, Jock D., and Shneiderman, Ben. "Readings in information visualization: using vision to think." *Morgan Kaufmann*, 1999.

[2] Shneiderman, Ben. "The eyes have it: A task by data type taxonomy for information visualizations." *Visual Languages, 1996. Proceedings., IEEE Symposium on*, 1996.

[3] Cui, Weiwei, et al. "Context preserving dynamic word cloud visualization." *Pacific Visualization Symposium (PacificVis), IEEE*, 2010.

[4] Viegas, Fernanda B., Martin Wattenberg, and Kushal Dave. "Studying cooperation and conflict between authors with history flow visualizations." *Proceedings of the SIGCHI conference on Human factors in computing systems. ACM*, 2004.

[5] Cliff, Kuang. "Infographic Of The Day: Using Twitter And Flickr Geotags To Map The World." http://www.fastcodesign.com/1664462/infographic-of-the-day-using-twitter-and-flickr-geotags-to-map-the-world. *Fast Codesign*, 12 July 2011. Web.

[6] Durdica, Sarjanovic. "Pogledajte Gdje Se Na Svijetu Najviše Koriste Flickr I Twitter." http://www.24sata.hr/tehnologija/pogledajte-gdje-se-na-svijetu-najvise-koriste-flickr-i-twitter-227828. *24sata*, 15 July 2011. Web.

[7] Available at http://blog.creativeaction.co.uk/2012/10/30/ibm-think-exhibit-data-wall/

[8] Rosling, Hans, Ronnlund A. Rosling, and Ola Rosling. "New software brings statistics beyond the eye." *Statistics, Knowledge and Policy: Key Indicators to Inform Decision Making. Paris, France: OECD Publishing*, 2005, 522-530.

[9] Chen, Xiang and Yang, Cheng-Zen. "Visualization of social networks." In:*Handbook of social network technologies and applications. Springer US*, 2010, 585-610.

[10] Hansen, Derek and Shneiderman, Ben. Smith, Marc A. "Analyzing social media networks with NodeXL: Insights from a connected world." *Morgan Kaufmann*, 2010.

[11] Brandes, Ulrik. "Social network analysis and visualization." *IEEE Signal Processing Magazine*, 2008, 25.6: 147-151.

[12] Lex, A., Gehlenborg, N., Strobelt, H., Vuillemot, R. and Pfister, H. 2014. "UpSet: Visualization of Intersecting Sets." *TVCG*, 2014.

[13] Shneiderman, Ben. "Designing the user interface-strategies for effective human-computer interaction." *Pearson Education India*, 1986.

[14] Fruchterman, Thomas MJ and Edward M. Reingold. "Graph drawing by force-directed placement." *Software: Practice and experience 21.11*, 1991, 1129-1164.

[15] Odewahn, Andrew. "Visualizing the US Senate social graph (1991-2009)." *Ch 8*, 2010, 123-142.

[16] Politiz, Available at http://www.politiz.org, 2012.

[17] Cox, Trevor F., and Michael AA Cox. "Multidimensional scaling." *CRC Press*, 2000.

[18] Doi, Kim et al. "Visualization of Movie Recommender System using Distribution Maps." *Proceedings of IEEE Pacific Visualization Conference (PacificVis'12), Songdo, South Korea*, February 2012.

[19] Kim, Gi-nam, et al. "Bubble Heap Graphs." *Proceedings of IEEE Information Visualization Conference (InfoVis' 13), Atlanta, USA*. 2013.

[20] Zanotti, Jim. "Open Knowledge Foundation: Europe's energy." Available at http://energy.publicdata.eu/ee/vis.html.

[21] Hoffman, Patrick, et al. "DNA visual and analytic data mining." *Visualization'97., Proceedings. IEEE*, 1997.

[22] Anton Strezhnev; Erik Voeten. "United Nations General Assembly VotingData."http://hdl.handle.net/1902.1/12379UNF:5:s7mORKL1ZZ6/P3AR5Fokkw== *Erik Voeten V7*. 2013.

[23] ISO 3166-1 alpha-3. Available at https://en.wikipedia.org/wiki/ISO_3166-1_alpha-3

[24] GovTrack.us: Tracking the U.S. Congress. Available at http://www.govtrack.us

[25] Zhu, Ying. "Introducing google chart tools and google maps API in data visualization courses." *IEEE computer graphics and applications* 6, 2012, 6-9.

[26] Bostock, Michael. "D3.js - Data Driven Documents.", 2012.

[27] Bostock, Michael. "Data-Driven Documents (d3.js), a visualization framework for internet browsers running JavaScript.", 2012.

[28] Joshua, Partlow. "List of 'Willing' U.S. Allies Shrinks Steadily inIraq." http://www.washingtonpost.com/wpdyn/content/article/2007/12/07/AR2007120702585.html. *The Washington Post*, 08 Dec. 2007. Web.

[29] Black, Ian. "Iran and Iraq Remember War That Cost More than a Million Lives." http://www.theguardian.com/world/2010/sep/23/iran-iraq-war-anniversary. *Theguardian*, 23 Sept. 2010. Web.

[30] Zanotti, Jim. "Israel: Background, Issues and U.S. Relations." http://www.fas.org/sgp/crs/mideast/RL33476.pdf. 8 Dec. 2014. Web.

[31] Lydia, Saad. "In the U.S., Blue States Outnumber Red States, 20 to 12." http://www.gallup.com/poll/160175/blue-states-outnumber-red-states.aspx *Gallup.com*, 30 Jan. 2013. Web.

[32] Chris, Cillizza. "The 9 Swing States of 2012." https://www.washingtonpost.com/blogs/the-fix/post/the-9-swing-states-of-2012/2012/04/16/gIQABuXaLT_blog.html. *The Washington Post*, 16 Apr. 2012. Web.

[33] Amy, Scullane. "Obama Wins Historic US Election." http://news.bbc.co.uk/2/hi/americas/us_elections_2008/7709978.stm. *BBC News*, 5 Nov. 2008. Web.

[34] Matthew, Liddy. "Results: 2012 US Presidential Election Map." http://www.abc.net.au/news/2012-09-21/2012-us-presidential-election-map-interactive/4272242. *ABC News*, 20 Sept. 2012. Web.

[35] Simon, Rogers. "2008 Presidential Election Results by State and County." http://www.theguardian.com/news/datablog/2009/mar/02/us-elections-2008. *Theguardian*, 9 Mar. 2009. Web.

[36] David, Jarman. "Comparing State-by-state Presidential Performance: 2012 vs. 2008." http://www.dailykos.com/story/2012/11/25/1164240/-Comparing-state-by-state-presidential-performance-2012-vs-2008. *Daily Kos*, 26 Nov. 2012. Web.

# Abstract

# N-polar Visualization : Interactive Visual Analytics with Multiple Interactive Anchors

Taeil Jeon

Program in Digital Contents and Information Studies

Department of Transdisciplinary Studies

The Graduate School

Seoul National University

Understanding the relationships between data objects is an important task that has received widespread attention in the field of visualization. Numerous efforts have been made to effectively visualize the relationships among individual data instances. The technique, however, is difficult to use for comparing the data objects against multiple target nodes in a single layout.

This study proposes a novel interactive visualization technique, named N-polar Visualization(NV). NV not only shows the general patterns among the data objects, but also it unravels that complex relationships between the data objects by placing the anchor nodes around the circle. Also, anchor nodes are movable on the circle, allowing users to interactively explore the

dataset. Users, therefore, can set the data instances of their interests as the anchor nodes, and observe how the rest of the nodes are related these anchor nodes and how they change by moving the positions of the anchors nodes on the circle.

Furthermore, a series of pilot studies are conducted to evaluate the feasibility of NV, using datasets where the relationships between the data objects are complicated. The studies could also demonstrate the generality of the proposed technique, by showing other such datasets can be examined using NV.

Finally, a pattern study and case studies are conducted to manifest the usefulness and effectiveness of NV. First, various types of patterns have emerged while examining the datasets using this technique, and they could be categorized into two groups: Linear Pattern and Clustering Pattern. Second, the feasibility of NV is further affirmed as it helped users to find unexpected insight during case studies based on the actual historical facts. The validity of these insights are confirmed by matching them with the statements articulated in news articles.

With NV, users can understand the relationships between data objects more effectively, and are expected to gain novel insights of the datasets, which are hard to attain with other visualization techniques.

**Keywords :** Data Object, Interactive Anchor, N-polar Visualization, Information Visualization, Visual Analytics

**Student Number : 2014-24822**