

점진적으로 확산하는 정보 네트워크의 시각화 도구 개발

유지연¹, 배준현², 김상욱³
경북대학교 전자전기컴퓨터학부
{jyyu¹, jhbae², swkim³}@cs.knu.ac.kr

Development of Visualization Tool for Information Diffusion Network

Jiyeon Yu¹, Joonhyun Bae², Sangwook Kim³
School of Electrical Engineering and Computer Science
Kyungpook National University

요 약

본 논문에서는 시간의 축에 따라 점진적으로 확산되는 정보 네트워크를 효과적으로 표현할 수 있는 시각화 도구를 소개하고자 한다. 이것은 기존의 정적인 데이터에 대한 정적인 시각화 결과물을 제시하는 시각화 도구들과는 달리, 시간의 축에 따라 동적으로 성장하는 데이터를 효과적으로 표현할 수 있다는데 그 차이가 있다. 이를 위해 그래프 기반의 정보 확산 네트워크 모델을 정의 하고, 이 그래프에 시간의 축에 따른 데이터의 성장을 반영하여, 정보 확산 과정을 시각화 할 수 있는 도구를 제작하였다. 결론적으로, 본 논문에서 제시하는 도구는 시간의 축에 따라 점진적으로 성장하는 정보 네트워크의 성장 과정을 시각화함으로써, 복잡하게 얽혀 있는 정보들이 어떻게 확산되며, 어떤 모양으로 생겼는지, 또 어떻게 진화하는가에 대한 시각적 재현을 통해 정보 확산 네트워크의 속성을 이해하는데 도움을 줄 수 있는 새로운 시각화 도구를 제시했다는데 그 의미가 있다.

1. 서론

복잡한 네트워크를 분석하고자 할 때 이들 간의 관계를 직관적으로 추론하고, 중요한 특징들을 쉽게 인지할 수 있도록 정교하게 시각화 하는 것은 중요한 문제가 된다 [1,2]. 이러한 시각화에서 그래프적 표현은 강력한 전달 효과와 명쾌한 비전 제시 효과의 장점을 가진다. 하지만 기존의 시각화 기법들은 대부분 정적인 데이터에 대해 정적인 결과물을 생성하는 형태가 대부분 이었다[3]. 이는 정보 확산 네트워크 처럼 시간의 흐름에 따라 점진적으로 성장해 가는 데이터를 효과적으로 시각화 하는 데는 한계가 있다.

본 논문에서는 정보 확산 네트워크와 같은 시간의 축에 따라 점진적 성장을 해가는 데이터를 효과적으로 시각화 할 수 있는 시각화 도구를 제시한다. 이것은 기존의 시각화 도구처럼 데이터의 순서나 계층을 표현하는 것이 아니라, 그래프의 성장 과정을 시각화 한다는데 차별을 둔다. 그 결과, 본 논문에서 제시한 그래프 성장 모델과 시각화 도구를 통해 정보가 어떤 모양으로, 어떻게 확산되어져 가는가에 대한 그 과정을 재현할 수 있게 되었다.

본 논문의 구성은 다음과 같다. 2절에서 기존의 네트워크 시각화 연구들을 소개하고, 3절에서는 본 논문에서

제안하는 정보 확산 네트워크 모델을 소개 한다. 4절에서는 제안된 모델을 적용한 시각화 도구의 구현환경과 구현결과를 소개하고, 마지막 5절에서 본 논문의 성과 제시한다.

2. 관련연구

그래프를 이용한 정보의 시각화 기법들은 오래전부터 연구가 되어져왔고 그를 이용한 많은 응용프로그램들이 개발이 되었다. 그래프를 이용한 시각화 응용에는 통신망 네트워크를 그래프[4], 지리 연결망의 네트워크[5], 의미들을 관계를 네트워크로 표현하는 시맨틱 네트워크[6], 하이퍼링크로 연결된 웹 구조 네트워크[7], 사람들 간의 사회적 관계를 나타낸 소셜 네트워크[8] 그리고 정보의 전파 과정을 표현하는 정보 확산 네트워크[9] 에도 사용이 되고 있다. 이러한 시각화 도구에서는 네트워크 전체를 표현하거나, 사용자가 관심을 가지고 있는 서브 그래프를 표현하는데 그 한계를 가진다. 일부 연구에서 다이나믹한 그래프 확장을 표현할 수 있는 방안을 제시 한 적이 있으나[10], 이런 동적 시각화 도구들 또한 데이터 통신량을 줄이거나 사용자의 심상 지도를 유지하며 정보를 표현하는데 주 안점을 두고 개발이 되어 왔다[11]. 이러한 기존 연구들은

정보 확산 네트워크의 성장 과정을 표현하는데 적합하지가 않다. 본 논문에서는 기존의 시각화 도구에서는 다루지 않았던 정보 확산 네트워크의 성장을 표현하는데 주안점을 두고 있다. 즉, 시간에 흐름에 따라 점진적으로 성장해 나가는 정보 확산 네트워크가 어떻게 성장을 해 왔는지, 어떻게 해서 이런 위상을 가지게 되었는지를 재현 할 수 있는 시각화 도구를 개발하고자 한다.

이를 위해 본 논문에서는 그래프 기반의 정보 확산 네트워크 모델을 정의 하고, 이 그래프에 시간의 축에 따른 데이터의 성장을 반영하여, 정보 확산 과정을 시각화 할 수 있는 도구를 제작하였다. 이제까지 정보 확산 과정을 표현하고자 한 시각화 도구는 제시된 바 없으며, 본 논문이 이러한 시각화의 첫 번째 시도라 믿고 있다.

3. 정보 확산 네트워크 시각화

본 절에서는 본 논문에서 제안하고자 하는 정보 확산 네트워크의 시각화를 위한 정보 확산 그래프 모델과 자료구조 그리고 시각화 알고리즘에 대해서 설명한다.

3.1 정보 확산 그래프 모델

먼저, $V_t = \{p_1, p_2, \dots, p_n\}$ 를 시간 t 에서 정보의 확산 네트워크를 구성하는 정보의 집합이라고 하자, 그리고 $E_t = \{(p, q) | p \in V, q \in V\}$ 를 시간 t 에서 확산에 대한 정보의 집합이라고 하자. 즉, 집합 V 에 속하는 어떤 정보 p 가 다른 정보 q 로 전파가 된다면, 집합 E 에 원소 (p, q) 가 추가된다. 여기서 정보의 전파는 시간에 흐름에 따라 발생함으로 집합 V 와 집합 E 는 시간의 축에 따라 성장하게 된다. 이를 시간 t 에서 그래프 $G_t = (V, E)$ 로 정의 한다. 이에 따라 정보 확산 네트워크는 이 정보 확산 그래프를 시간 t 에 따라 성장시켜 더 이상 성장 할 데이터가 없을 때 성장이 종료되게 된다.

3.2 정보 확산 그래프 자료구조

본 연구에서 시간의 축에 따라 성장하는 그래프를 표현하기 위해 그래프 데이터 와 그래프 뷰 라는 두 개의 자료구조를 정의 하였다.

이것은 기존의 시각화 도구에서와 달리, 그래프에 존재하는 노드와 간선의 관계가 단순한 연결 관계가 아닌 시

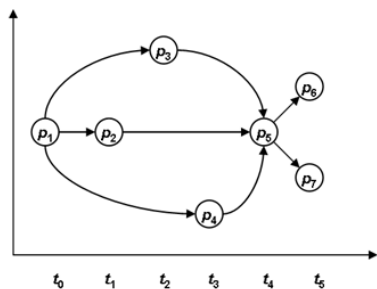


그림 1 시간에 따른 정보 확산 그래프의 성장 모형

간의 축에 따라 표현(representation)이 되어야할 관계이기 때문이다. 만일 전통적인 그래프 자료구조를 정보 확산 그래프에 적용 시킬 경우 시간의 축을 무시하는 표현이 발생하게 된다. 이러한 이유로 본 논문에서는 그래프의 데이터와 뷰를 분리 하였고, 이를 통해 그래프 성장 과정이 시간의 축을 벗어나지 않으며, 점진적인 그래프 성장을 통해 효율적인 데이터 트랜잭션이 이루어 질 수 있도록 하였다. 여기에서 데이터 트랜잭션은 특정 정보의 확산 정보를 DB로부터 획득하는 것을 말한다. 그리고 노드의 확장은 이 정보(노드)가 다른 정보(노드)에게로 확산 된 데이터가 존재할 때, 이 확산 데이터를 획득하는 것을 말한다. 그래프 데이터와 그래프 뷰의 정의는 다음과 같다.

그래프 데이터 D 는 t 시간에 시각화 도구가 가지고 있는 모든 정점과 간선 정보를 말한다. 특정 시간 t 에서의 데이터 D_t 는 확장을 마친 노드의 집합 V_p 와 확장이 되지 않은 노드의 집합 V_q 그리고 이들 노드간의 연결 관계 E 로 표현 할 수 있다. 여기서, $V = V_p \cup V_q$ 이고 $V_p \cap V_q = \emptyset$ 이다. 그래프 뷰는 t 시간에 화면상에 표현되는 그래프를 말한다. 그래프 데이터에서 정의한 D 는 시간 t 를 기준으로 t 에 포함되어 화면에 표현 된 노드의 집합 S_{show} 와 아직 t 에 포함 되지 않아 표현 되지 않은 노드 집합 S_{hidden} 로 나뉜다. 즉, $V = S_{show} \cup S_{hidden}$ 이고 $S_{show} \cap S_{hidden} = \emptyset$ 이다.

3.3 정보 확산 그래프의 성장과 시각화 과정

정보 확산 그래프의 성장은 데이터의 성장과 뷰의 성장을 통해 이루어진다. 표현할 데이터를 DB로부터 점진적으로 획득해 나가는 것이 데이터의 성장이고, 획득된 데이터를 시간 t 에 따라 점진적으로 표현해 나가는 것이 뷰의 성장이다.

데이터의 성장은 V_q 집합의 한 노드가 V_p 로 전이 되며 이루어진다. 이 전이는 데이터 획득의 관점에서 볼 때 노드의 확장과 그 의미가 같다. 노드가 확장을 하면, 이 노드로부터 파생된 노드들과 이들 간의 간선 정보를 다시 데이터에 삽입한다. 이 과정을 반복하여 V_q 의 모든 노드가 V_p 로 전이 되면 데이터의 성장이 종료 된다. 뷰의 성장은 시간 t 에 따른 D_t 에서 $S_{visible}$ 를 추출하여 이를 그래프로 시각화 하며 이루어진다. 그리고 더 이상 시각화할 데이터가 없을 때, 즉 S_{hidden} 이 공집합이 되면 그래프의 모든 성장이 종료 된다.

그림 2는 그림 1의 정보 확산 그래프를 시각화 하는 과정을 나타낸 그림이다. 2(a) 와 2(b) 는 각각 데이터의 성장 과정과 뷰의 성장 과정을 나타내고 있다. 여기에서 그래프의 성장은 6번에 걸쳐 이루어졌는데, 최초 초기화 단계에서 V_q 와 S_{hidden} 에 각각 루트 노드인 p_1 이 삽입 된다. 초기화 단계가 지나 t_0 가 되면 초기화 과정에서 삽입된 V_q 의 p_1 이 확장을 하게 되고, 이를 통해 p_2, p_3, p_4 가 V_q 와 S_{hidden} 에 삽입이 된다. 같은 iteration 에서 S_{hidden} 에 존재하는 가장 오래된 노드의 시간을 찾아 이

(a) 데이터 성장 과정

(b) 뷰 성장 과정

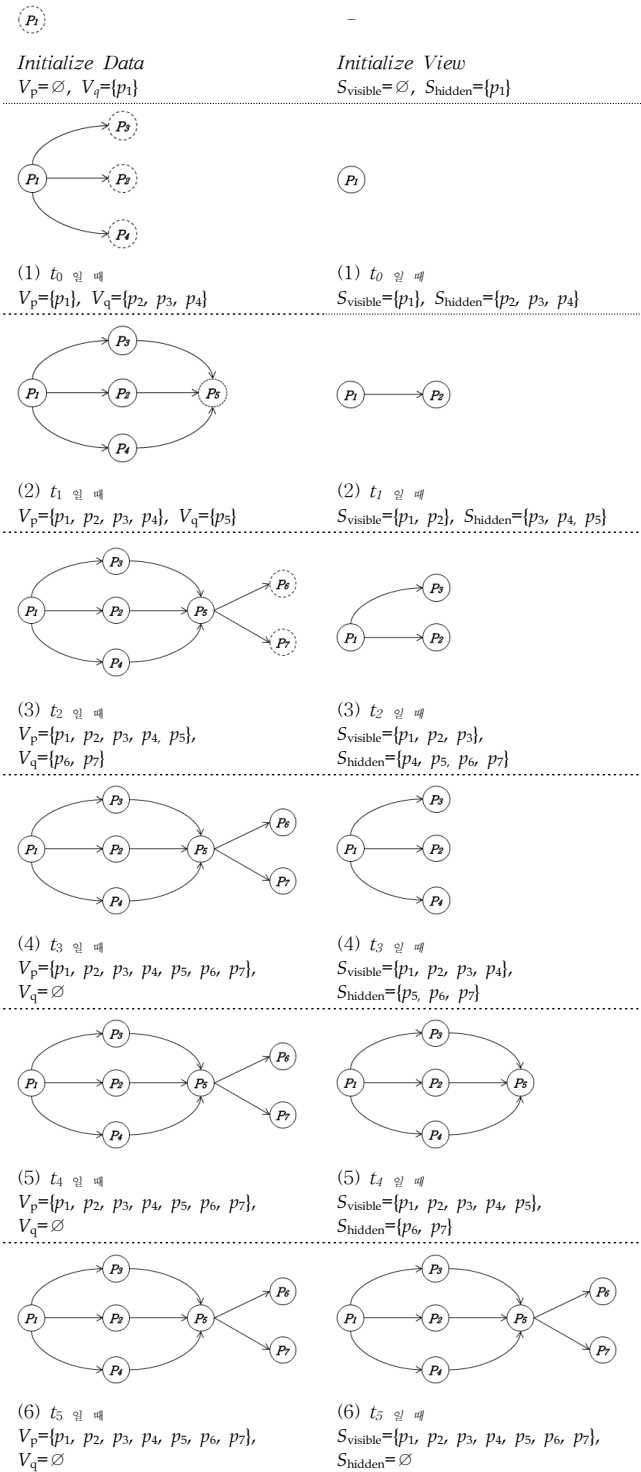


그림 2 그래프 시각화 과정

시간에 해당하는 모든 노드들을 $S_{visible}$ 즉, 그래프로 표현을 한다. 여기에서는 p_1 이 S_{hidden} 에서 가장 오래된 노드이므로 p_1 이 그래프 상에 표현이 된다. t_1 이 되면 V_q 에 있던 p_2, p_3, p_4 가 확장을 하여 p_5 가 V_q 와 S_{hidden} 에 삽입 된다. 그리고 S_{hidden} 에 존재 하는 가장 오래된 노드인 p_2 가 그래프에 표현이 된다. 이 과정을 반복하여 S_{hidden} 에 더 이상 노드가 없게 되면 그래프의 모든 성장이 종료

표 1 정보 확산 그래프 성장 의사코드

```

Input the root node root
Initialize expandable, hidden to empty queue
Initialize force directed placement layout gFDL
Push root to expandable and hidden
Do
  // expand phase
  for each node of expandable
    node.expand()

  // find oldest node from hidden
  oldestNodes = popOldestNodes( hidden )
  for each this_node of oldestNodes
    gFDL.add( this_node ) // add node to graph
    // find edges that destination is this_node and
    // linking each other
    gFDL.linkage( this_node ) // linking
    Push node to hidden
  gFDL.draw() // draw graph
While hidden is not null
    
```

되게 된다. 표 1은 제안된 정보 확산 그래프의 성장 알고리즘 의사코드이다.

4. 구현 및 결과

본 절에서는 정보 확산 그래프 모델을 적용하여 개발한 시각화 시스템과 구현 결과를 소개한다.

4.1 시스템 구현

본 연구에서는 그래프의 동적인 성장을 위해 온라인 기반의 시각화 시스템을 구성 하였다. 온라인 기반의 시각화 시스템은 시각화해야 할 데이터가 매우 거대하여 전체를 한 번에 프로세싱하기 힘들거나, 빠른 속도로 서브 그래프를 추출 할 때나, 사용자의 인터랙션에 의해 시각화가 이루어져야 할 때 주로 사용이 되어왔고[3], 본 연구에서도 유사한 이유로 채택을 하였다. 본 논문에서 개발한 시각화 도구는 ACM Digital Library 로부터 논문의 인용 정보를 수집하는 Java 기반의 크롤러, 데이터 저장을 위한 MySQL DBMS, 뷰어에서 요청한 정보를 DB 로부터 가져와 XML 형태로 반환해 주는 PHP 로 제작된 WAS, 그리고 정보 확산 그래프를 시각화 하는 FLEX 로 개발된 뷰어로 구성된다. 뷰어에서는 그래프 드로잉을 위해 Force Directed Layout[2] 을 사용하였다.

4.2 구현 결과

그림 3 는 본 논문에서 시각화 시스템의 시각화 과정을 캡처한 것이다. 여기서 각 정점들은 논문이 되고 간선들은 인용을 나타낸다. 이 뷰어에서 색상과 크기라는 시각화의 장점을 이용하여 ACM 디지털 라이브러리의 분류체계 중 가장 상위 카테고리들 각기 다른 색상으로 표시하고 논문의 인용 횟수에 따라 크기를 다르게 하여 네트워크 속성의 직관적 이해를 도왔다. 또한 시간의 흐름에 따른 네트워크 형성 과정을 시각화함으로써 해서 정보들이 어떻게 확

참고문헌

[1] I. Herman, G. Melancon, and M. S. Marshall, "Graph Visualization and Navigation in Information Visualization: A Survey", IEEE Trans. Visualization and Computer Graphics, vol. 6, no. 1, pp. 24-43, 2000.

[2] T. M. J. Fruchterman, E. M. Reingold, "Graph drawing by force-directed placement", Software - Practice & Experience, v.21 n.11, p.1129-1164, Nov. 1991

[3] G. D. Battista, P. Eades, R. Tamassia, and I. G. Tollis, "Graph Drawing: Algorithms for the Visualization of Graphs", Prentice Hall PTR, Upper Saddle River, NJ, 1998

[4] T. He, "Internet - Based Front - End to Network Simulator", Data Visualization '99, Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization, Springer - Verlag, pp. 247 - 252, 1999.

[5] U. Brandes, G. Shubina, and R. Tamassia, "Improving Angular Resolution in Visualizations of Geographic Networks", to be published in: Data Visualization '2000, Proceedings of the Joint Eurographics and IEEE TCVG Symposium on Visualization, Springer - Verlag, 2000.

[6] D. J. Duke, K. W. Brodlie, D. A. Duce, and I. Herman, "Do You See What I Mean?" IEEE Computer Graphics and Applications, vol. 25, no. 3, May/June 2005.

[7] K. Andrews, "Visualizing Cyberspace: Information Visualization in the Harmony Internet Browser", Proceedings of the IEEE Symposium on Information Visualization (InfoViz'95), IEEE CS Press, pp. 97 - 105, 1995.

[8] J. Blythe, C. McGrah, and D. Krackhardt, "The Effect of Graph Layout on Inference from Social Network Data", Proceedings of the Symposium on Graph Drawing, GD '95, Springer - Verlag, pp. 40 - 51, 1995.

[9] Y. Takama, A. Matsumura, and T. Kajinami, "Visualization of News Distribution in Blog Space" Proceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology, p.413-416, December 18-22, 2006

[10] M. L. Huang, P. Eades, and J. Wang, "Online Animated Graph Drawing Using a Modified Spring Algorithm", Journal of Visual Languages and Computing, Vol. 9 No. 6, 1998.

[11] K. Misue, P. Eades, W. Lai, and K. Sugiyama, "Layout Adjustment and the Mental Map", Journal of Visual Languages and Computing, Vol. 6, pp. 183 - 210, 1995.

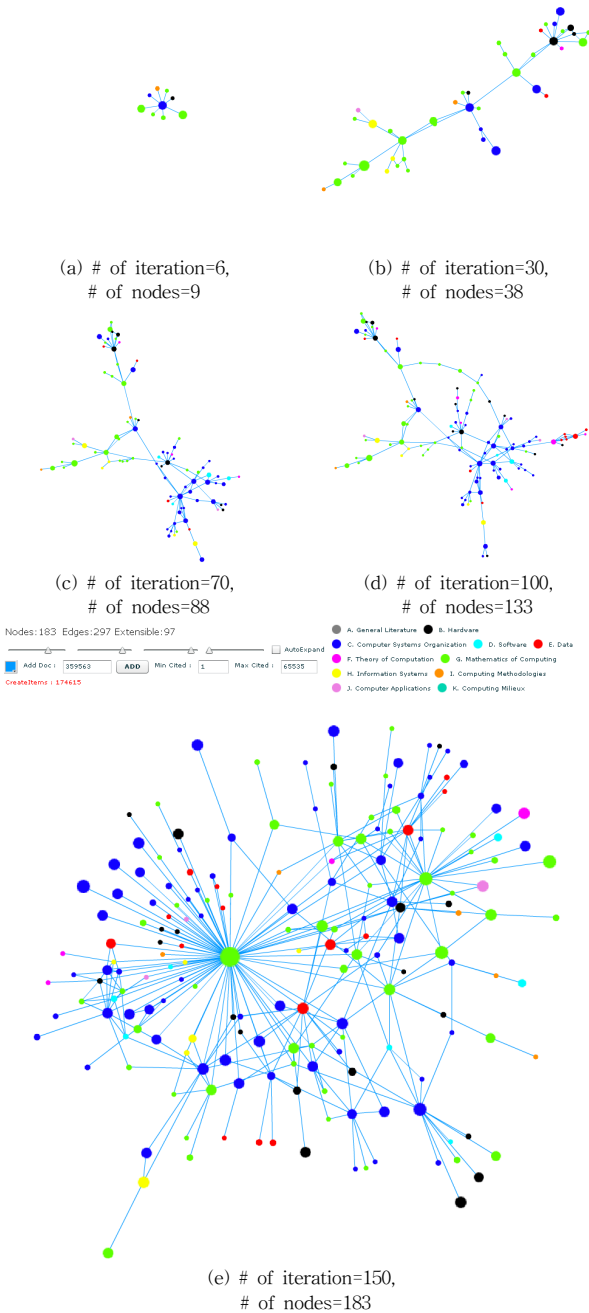


그림 3 논문 인용 정보의 시각화 결과

산되며, 어떤 모양으로 생겼는지, 또 어떻게 진화하는가에 대해 시각적으로 표현할 수 있게 되었다.

5. 결론

본 논문에서 우리는 시간에 따라 점진적으로 성장하는 정보 확산 네트워크의 시각화 도구를 제시하고, 논문의 인용 그래프를 통해 제안하는 모델에 대한 실증적 실험의 결과를 제시 하였다. 본 논문의 주요 공헌으로는 시간의 축이 반영된 정보 확산 네트워크의 그래프 구조를 모델링 하고, 이를 통해 시각화 도구를 구현하여 정보 확산 과정의 시각적 재현을 가능케 했다는데 있다.