ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2015 September: 21(3): 19~36 http://dx.doi.org/10.13088/jiis.2015.21.3.19

일반영향요인과 댓글기반 콘텐츠 네트워크 분석을 통합한 유튜브(Youtube)상의 콘텐츠 확산 영향요인 연구*

박병언
한양대학교 경영대학
(pbe@hanyang.ac.kr)

임규건
한양대학교 경영대학
(gglim@hanyang.ac.kr)

대표적 소셜미디어인 유튜브는 기존 폐쇄형 콘텐츠 서비스와는 다르게 개방형 콘텐츠 서비스로 이용자들의 참여와 공유를 통하여 많은 인기를 유지하고 있다. 콘텐츠 산업에서 중요한 위치를 차지하고 있는 유투브 상의 콘텐츠 확산 요인에 관한 기존의 연구들은 댓글 수 등과 같은 일반적 정보 특성 요인과 조회 수 간에 상관관계 등을 분석하는 것이 대부분이었다. 최근 네트워크 구조를 기반으로 한 연구들도 진행되었으나 대부분 콘텐츠를 이용하는 대상인 구독자나 지인 등을 중심으로 한 인적 관계 네트워크 구조 연구가 대부분이었다. 이에 본 연구에서는 실질적인 콘텐츠를 중심으로 한 네트워크 구조와 일반요인을 통합한 모델을 제시하고 확산요인을 분석하고자 한다. 이를 위해 통합 모델 인과관계 분석과 함께 21,307개의 유튜브 콘텐츠를 콘텐츠 기반 네트워크 구조로 분석하였다. 본 연구를 통해 기존에 알려진 일반적 요인과 네트워크 요인들이 모두 조회수에 영향을 주는 인과관계를 통계적으로 재검증하였으며 통합적으로는 등록자의 구독자 수, 경과시간, 매개 중심성, 댓글 수, 근접 중심성, 클러스터링 계수, 평균 평점 순으로 조회 수에 긍정적인 영향을 미치는 것으로 분석되었다. 하지만 네트워크 요인중 연결정도 중심성과 고유벡터 중심성은 부정적 영향을 주는 것으로 분석되었다. 본 연구를 통하여 유튜브 콘텐츠 확산에 대한 일반영향요인과 구조적인 현상을 함께 규명하였다. 본 연구는 기업들이 유튜브와 같은 콘텐츠 서비스를 통한 온라인 마케팅 활동 시 콘텐츠들의 구조적인 면을 고려할 수 있는 근거를 제공하였으며 음반산업의 수요예측이나 콘텐츠 제작 업체들의 원활한 서비스 제공을 위한 설명력있는 영향요인 및 모델이 될 수 있을 것이다.

주제어 : 콘텐츠 네트워크, 유투브, 확산 요인, 소셜미디어, 영향 요인

논문접수일 : 2015년 6월 11일 논문수정일 : 2015년 9월 7일 게재확정일 : 2015년 9월 9일
투고유형 : 국문일반 교신저자 : 임규건

1. 서론

소셜 미디어(Social Media)가 등장하기 시작하면서 개인의 참여와 공유를 유도하는 서비스들이 확산되고 있다. 기존에 폐쇄적으로 운영되던 콘텐츠 및 미디어 서비스들과는 다르게 최근의 서비스들은 이용자들 간에 상호작용을 하며 콘텐츠의 생산에서 소비까지 자생적으로 이루어지고 있다 (Kim et al., 2014; Cho and Kim, 2011; Lee et al., 2009). 이러한 시대 흐름에 따라 소셜

* 이 논문은 박병언 석사학위논문을 기반으로 2015년 지능정보시스템학회 춘계학술대회에서 초록이 발표되어 수정 보완되었음(Park and Lim, 2015). 이 논문은 한양대학교 교내연구지원사업으로 연구되었음(HY-2013년도).

미디어 기반인 유튜브(Youtube)는 동영상 콘텐츠를 제공하는 온라인 오픈 플랫폼(Open Platform)으로 구글(Google)에 이어 세계 제2위의 정보검색 서비스로 월 8억 명에 달하는 전세계 사용자가 이용하고 있다. 유투브에서는 2012년을 기준으로 하루 40억 개 이상의 동영상이 재생되는 것으로 집계되고 있다.

또한, 유튜브는 일반 콘텐츠 산업 뿐만 아니라 음반 산업에 대해서도 그 영향력이 커지고 있다. 빌보드 차트(Billboard Charts)에서는 2013년부터 온라인 스트림(Online Stream) 분야를 새롭게 포함하고 있으며, 이 밖에 여러 음악 차트에서도 유튜브의 순위방법론이 활용되고 있다. 이와함께 유튜브의 조회 수 등은 음반 산업에서 수요예측이나 다양한 마케팅 활동의 성과지표로 활용되고 있다. 유튜브와 관련한 주요 연구주제로는 크게 첫째, 기존 폐쇄형 콘텐츠 서비스와는 다른 일반 이용자도 콘텐츠를 제공하는 개방형 콘텐츠 서비스가 가지는 특성 및 현상에 대한 연구주제가 있다. 둘째, 어떠한 콘텐츠가 인기가 있으며 콘텐츠가 어떻게 확산되고 유통되는 지에 대한 활용성을 높이기 위한 방안 및 추천 서비스 등의 연구가 있다.

유튜브의 콘텐츠 확산과 관련된 연구를 살펴보면 Chatzopoulou et al.(2010)의 경우 댓글 수 등과 같이 일반적인 요인과 조회 수와의 상관관계에 대해 연구하였다. 해당 콘텐츠를 즐겨찾기에 추가하거나 댓글을 다는 행위로 관련 동영상들의 방향성 네트워크를 구축하여 조회 수와의 관계를 연구하였다. 하지만 상관분석을 통한 결과는 간접적이며 명확한 결과를 도출하지 못하였다. 이후에도 소셜 미디어인 유튜브의 특성으로 인해 콘텐츠 확산과 네트워크 구조 간의 관계에 대해 많은 연구가 있었다. Yoganarasimhan(2012)는 기존 일반요인에 세분화한 친구와 구독자간의 네트워크 요인을 추가하여 조회 수에 미치는 영향을 시점별로 분석하였다. 그리고 Susarla(2012)는 네트워크가 없는 일반적인 경우와 친구 그리고 구독자를 기반으로 한 네트워크를 고려한 경우의 차이를 콘텐츠 경과시간에 따라 분석하였다.

이러한 연구들은 댓글 수 등과 같이 기존에 널리 알려진 일반적인 정보적 특성 또는 요인과 조회 수 간에 상관관계 등을 분석하는 정도에 머물렀다. 또는 콘텐츠를 이용하는 대상인 구독자나 친구 등을 중심으로 한 네트워크 연구가 대부분이었다. 즉, 실질적인 콘텐츠를 중심으로 한 네트워크를 이용하여 콘텐츠들의 구조적인 측면을 고려하지 못한 점이 있었다.

이에 본 연구에서는 유튜브(Youtube) 콘텐츠의 실질적인 확산을 나타내는 지표로 사람들이 콘텐츠를 본 양적 데이터인 조회 수로 정의하고 이에 대한 영향요인분석을 일반적 요인과 네트워크의 구조적 요인을 함께 고려하여 분석하는 통합 모델을 제시하고자 한다. 이를 위해 기존에 연구된 일반요인들과 이용자 참여 등을 통해 형성되는 콘텐츠의 댓글기반 네트워크를 이용하여 콘텐츠들 간의 구조적인 측면이 콘텐츠 확산에 어떠한 영향을 주는지 분석하고 규명하고자 한다.

2. 선행연구

2.1 소셜 미디어에 대한 연구

소셜미디어란 웹 기술을 기반으로 이용자 제작 콘텐츠(UGC)등의 생산과 공유를 가능하게

하는 플랫폼으로 정의된다(Kaplan and Haenlein, 2010). 이러한 소셜 미디어는 자유롭게 콘텐츠나 정보를 공유하며 참여할 수 있는 오픈 서비스(Open Service)의 형태이다. 오픈 서비스의 가장 대표적인 특징은 소비자가 생산자의 역할까지 하는 프로슈머(Prosumer)에 기반을 두고 있다는 점이다. 즉, 오픈서비스의 대표적인 동영상 공유 웹사이트인 유튜브는 이용자가 스스로 방송 콘텐츠를 제작하고 공유하는데 초점이 맞춰져 있다(Burgess and Green, 2009).

이용자가 참여함으로 인해서 유튜브는 동영상 콘텐츠 공유 웹 사이트임과 동시에 대표적인 온라인 소셜 네트워크 서비스의 하나로 꼽히고 있다. 또한 콘텐츠 중심의 네트워크 모델(Content Oriented Network Model) 방식 또는 콘텐츠 중심의 OSNs(Online Social Networks) 등으로 정의되고 있다(Pallis etal., 2011).

따라서 소셜미디어인 유튜브는 일반적인 콘텐츠 서비스와는 다르게 사용자들의 공유와 참여로 많은 데이터가 축적되어있다고 판단할 수 있다. 또한 이렇게 축적된 데이터들을 활용하여 형성되는 네트워크 구조와 조회 수 간에 관계에 대해 연구할 필요성이 있다.

2.2 네트워크 구조에 대한 연구

사회네트워크 분석에서 가장 기초적인 단위로 노드(node)와 연결선(edge)이 있으며 이를 통해 형성되는 네트워크 구조는 '방향 네트워크(directed network)'와 '무방향 네트워크(undirected network)'로 구분할 수 있다. 방향 네트워크는 Susarla (2012)의 연구와 같이 친구 맺기나 구독하기 등과 같이 직접적인 행위를 통하여 방향성을 가진다. 방향 네트워크에 대한 연구는 대부분 이용자나 사람과 같이 실질적인 행위를 할 수 있는 대상을 중심으로 이루어졌다(Chatzopoulou et al., 2010; Yoganarasimhan, 2012). 무방향 네트워크는 방향성을 가지지 않으며 간접적이거나 양방향적인 관계를 가진다. 방향 네트워크처럼 정확한 관계를 표현하기는 힘들지만 행위를 할 수 없는 콘텐츠와 같은 대상으로도 네트워크 구조를 형성할 수 있다는 장점이 있다. 이러한 간접적인 네트워크의 구축은 실질적인 콘텐츠 이용 시 사용하는 검색엔진 및 키워드를 이용하거나 이용자들이 콘텐츠를 보고 남기는 정보들의 특정 패턴 등을 통하여 이루어진다. 본 연구와 같이 실질적인 행위가 이루어지지 않는 콘텐츠들의 구조적인 측면에 대한 연구는 콘텐츠를 중심으로 한 무방향 네트워크가 가장 적합함을 알 수 있다.

지금까지 유튜브를 네트워크 구조화 하기 위해 많은 연구가 이루어져 왔다. 이러한 네트워크 구조화 방식을 Santos et al.(2007)는 4가지로 정리 및 제시하였으며 관련 연구들은 아래 <Table 1>과 같다.

〈Table 1〉 Classification of Network Structure Methods for Youtube

Classification	Researchers
User-User friendship	Susarla(2012), Yoganarasimhan(2012)
User-User subscription	Susarla(2012), Yoganarasimhan(2012)
User-Video favoring	Chatzopoulou et al.(2010),
Video-Video relatedness	Akroufet al.(2013)

유튜브의 순수한 인적 네트워크 구조화는 친구 맺기나 특정 이용자를 정기 구독하는 행위로 구축된다. 그리고 콘텐츠가 연계된 경우는 해당

콘텐츠를 특정 이용자가 즐겨찾기에 추가하는 등의 행위로 이루어진다. 마지막으로 순수한 콘텐츠 네트워크는 유튜브의 검색엔진을 통하여 이루어진다.

하지만 이렇게 단순히 검색엔진만을 이용한 콘텐츠 네트워크는 많은 한계점을 가지며 네트워크 구조를 세밀하게 형성 및 구조화하기 힘들다는 단점이 있다. 따라서 본 연구에서는 콘텐츠 간 네트워크 구축 시 동일 댓글자 등 추가적인 방안을 통해 보다 세밀한 네트워크 구축을 진행하였다.

2.3 네트워크 요인에 대한 연구

네트워크상에서 특정 노드의 구조적인 측면을 측정하기 위한 대표적인 네트워크 요인 및 척도로 중심성(centrality)이 있다(Wasserman and Faust, 1994). 노드의 중심성은 전체 네트워크에서 중심에 위치하는 정도를 표현하는 지표로 정의할 수 있다. 중심성(centrality)은 관점에 따라 구분되며 가장 대표적인 지표는 크게 3가지로 연결정도 중심성(degree centrality), 근접 중심성(closeness centrality), 매개 중심성(betweenness centrality)이 있다(Freeman, 1979; Bolland, 1988; Wasserman and Faust, 1994; Costenbader and Valente, 2003).

연결정도 중심성은 네트워크에서 얼마나 많은 노드들이 연결되어 있는지를 나타낸다. 한 개의 노드가 많은 연결을 가질수록 많은 기회와 영향력을 가진다고 보는 것이다. 근접 중심성은 한 노드가 다른 노드들에 얼마나 가까이 위치해 있는가를 나타내는 개념이다. 다른 노드들과 가깝게 위치할수록 거리적인 이점을 가지어 관계를 가질 때 더 쉽다고 볼 수 있다. 매개중심성은 다른 노드들을 연결해주는 중개 또는 매개 역할을 할 수 있는 정도를 나타낸다. 특정노드는 매개자 또는 중개자 역할을 수행하는 노드를 거쳐야하므로 매개중심성이 높은 노드는 정보의 흐름에 크게 관여하는 것으로 알려져 있다.

여기에 근처 노드의 위상 혹은 인기도에 영향을 받는 정도를 고려하여 고유벡터 중심성이 추가되었다(Bonacich, 1987). 고유벡터 중심성은 네트워크 구조에 관계한 노드들의 영향력을 고려하여 해당 노드의 위세를 나타낸다. 즉, 중심성이 낮은 노드보다 높은 노드들에 연결된 노드일수록 높은 영향력을 가진다고 보는 것이다. 실제로 구글(Google)의 Brin and Page(1998)는 고유벡터 중심성을 이용하여 검색엔진에서 순위를 정하는 PageRank 알고리즘을 개발하였으며 실제로 사용하고 있다.

마지막으로 위와 같이 네트워크 내 위치에 기반한 중심성이 아닌 특정 노드가 관계한 노드들의 연결이나 군집정도에 따라 산출되는 클러스터링 계수(Clustering Coefficient)가 있다. 클러스터링 계수는 한 노드에 연결된 임의의 두 노드가 서로 연결되어있을 확률로 군집이 형성되어 있는 경우에 나타나게 된다. 특정 노드에 연결되어 있는 노드들 간에 관계가 많이 형성되어 있을 경우 높은 것으로 계산된다. 즉, 네트워크 내에서 군집이 형성되는 경우 연결된 노드들 간에 연계가 활발하여 정보공유 등에 긍정적인 효과 등을 가져올 수 있다.

따라서 네트워크 구조의 다양한 측면들을 알아보기 위하여 대부분의 중심성 요인들과 클러스터링 계수를 사용할 필요가 있다.

2.4 유튜브(Youtube)의 콘텐츠 확산에 대한 연구

유튜브에서 콘텐츠의 인기나 확산을 대표할 수 있는 요인은 조회 수로 볼 수 있다. 이런 유튜브 콘텐츠의 조회 수에 영향을 미치는 요인에 대해 많은 연구가 이루어져왔다.

Chatzopoulou et al. (2010)는 유튜브의 조회 수에 대한 일반적인 영향요인으로 댓글 수, 즐겨찾기 등록 수, 좋아요, 평균평점으로 정의 후에 상관관계를 분석하였다. 댓글 수, 즐겨찾기 등록 수, 좋아요는 조회 수와 높은 상관관계를 가지는 것으로 나타났으나 평균평점의 경우는 약한 상관관계를 나타내었다. 또한 콘텐츠 등록 후 경과시간을 기준으로 각 요인들의 변화를 그래프로 확인하였을 때 조회 수와는 반대로 낮아지는 것으로 나타났다. 그리고 추가적으로 해당 콘텐츠를 즐겨찾기에 추가하거나 댓글을 다는 행위로 유사 동영상들의 방향성 네트워크를 구성하고 네트워크 요인인 내향중심성과 조회 수와의 관계를 연구하였다. 가장 상위권에 속해있는 콘텐츠들은 많은 네트워크를 구성하고 있는 것으로 나타났으나 조회 수와 상관관계가 강하지 않은 것으로 나타났다.

이러한 연구는 조회 수와의 관련성이 존재하나 그 관계를 명확하게 규명하지 않았으며 추가적인 연구가 필요했다. 그 후에도 유튜브 조회 수에 영향을 미치는 요인으로 일반적인 요인과 친구 또는 구독자 네트워크와의 관계에 대해서 많은 추가 연구들이 진행되었다.

Yoganarasimhan(2012)의 경우는 일반적인 요인과 네트워크 요인이 조회 수에 미치는 영향을 시점별로 분석하였다. 일반적인 요인으로 댓글 수, 좋아요, 평균평점, 즐겨찾기 등록 수, 콘텐츠 명성으로 정의하고 친구와 구독자 네트워크를 구성하고 연결정도와 매개중심성 그리고 클러스터링 계수를 추가하여 연구를 진행하였다. 콘텐츠를 등록한 시점이 이른 경우 댓글 수와 즐겨찾기 등록 수는 조회 수에 영향력을 미치는 것으로 나타났지만 평균평점과 콘텐츠의 명성은 어느 시점에서도 영향을 미치지 않았다. 그리고 콘텐츠 등록시점이 이른 경우 연결정도가 영향을 미치며 그 후에는 친구간의 네트워크나 매개중심성 그리고 클러스터링 계수 등이 영향을 미치는 것으로 분석되었다.

Susarla(2012)는 조회 수 확산에 대해 일반적인 요인과 구독자 및 친구에 대한 네트워크 요인이 단순히 영향을 미치는 것에 그치는 것이 아니라 콘텐츠 등록 후 경과시간에 따라 나타나는 네트워크 효과의 차이도 고려하여 연구하였다. 일반적인 요인으로는 콘텐츠의 평균평점, 경과시간, 다른 매체로의 링크 수로 정의하였고 네트워크 요인의 경우 구독자와 친구 네트워크를 구분한 연결정도 중심성을 사용하였다. 그 결과 콘텐츠의 경과시간을 고려하지 않고 전체적으로 봤을 때 평균평점과 링크 수 그리고 구독자 네트워크는 조회 수에 긍정적인 영향을 미치는 것으로 나타났다. 하지만 경과시간이나 친구 네트워크는 부정적인 영향을 미치는 것으로 나타났다. 추가적으로 경과시간을 고려하여 분석하였을 때 콘텐츠 등록 초기의 경우 구독자 네트워크가 더 큰 영향력을 나타내며 시간이 흐름에 따라 친구 네트워크가 상대적으로 더 많은 영향을 나타내는 것으로 분석되었다.

즉, 앞선 연구들로 인하여 유튜브의 조회 수 및 확산에 일반적인 요인 뿐 아니라 네트워크 구조적인 측면도 영향을 미치는 것을 알 수 있다. 하지만 대부분 친구 또는 구독자 네트워크 등 사

람을 중심으로 한 네트워크 연구가 많았다.

따라서 본 연구에서는 사람과 같이 직접적인 행위가 이루어지며 연결 및 형성되는 이용자 중심 네트워크가 아닌 직접적인 행위가 이루어지지 않는 콘텐츠 중심의 네트워크를 사용하여 조회 수와의 직접적인 인과관계에 대해 연구하고자 한다.

3. 연구문제와 모형

3.1 연구문제

본 연구에서는 일반적인 요인과 네트워크 요인이 콘텐츠의 확산을 나타내는 조회 수에 어떠한 영향을 미치는지 알아보고 실증적인 검증을 하고자 하였으며 연구문제는 다음과 같다.

연구문제 1 : 콘텐츠의 일반적인 정보적 특성과 조회 수 간에 인과관계가 있을 것인가?

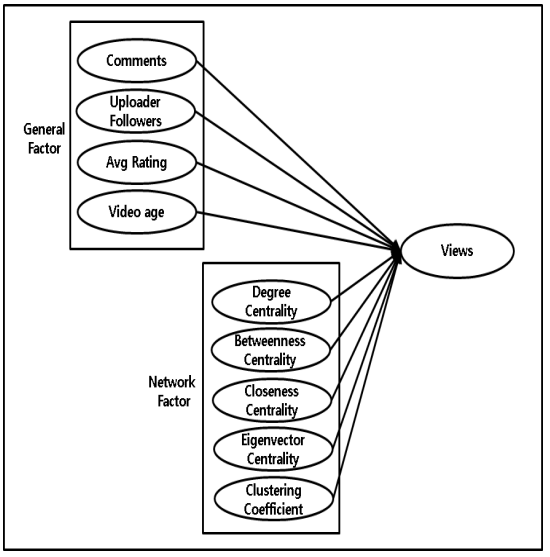
기존 연구에 의하면 콘텐츠의 일반적인 요인(조회 수, 댓글 수, 등록자의 구독자 수, 경과시간) 등은 조회 수에 정(+)의 영향을 미치는 것으로 파악된다. 이에 대한 검증을 실시한다. 또한 어떤 요인들이 더 영향을 크게 주는지 검증한다.

연구문제 2 : 콘텐츠의 네트워크 구조와 조회 수 간에 인과관계가 있을 것인가?

네트워크 요인(연결정도 중심성, 매개중심성, 근접 중심성, 고유벡터 중심성, 클러스터링 계수) 등이 조회 수에 정(+)의 영향을 미치는지 검증한다. 또한 어떤 요인들이 더 영향을 주는지 검증한다.

연구문제 3 : 일반요인과 네트워크 구조 영향요인 상의 차이가 있을 것인가?

일반적인 요인(조회 수, 댓글 수, 등록자의 구독자 수, 경과시간)과 네트워크 요인(연결정도 중심성, 매개중심성, 근접 중심성, 고유벡터 중심성, 클러스터링 계수)의 통합 모형에서 조회수에 영향이 어떻게 다르게 나타나는지 검증한다. 이를 통해서 어떤 요인들이 더 큰 영향을 주는 요인인지 검증한다.



〈Figure 1〉 Research Model

3.2 연구모형

위의 연구 문제를 검증하기 위하여 본 연구의 연구모형은 <Figure 1>와 같이 크게 '일반 요인'과 '네트워크 요인 ′ 으로 구성하여 조회수에 미치는 영향을 연구하였다. 일반 요인은 기존 연구에서 연구되었던 콘텐츠에 대한 댓글 수, 등록자의 구독자 수, 평점, 경과시간으로 구성되어 조회 수와의 인과관계를 검증한다. 그리고 콘텐츠

들의 구조적인 측면과 조회 수와의 관계를 알아보기 위하여 추가한 네트워크 요인은 연결정도 중심성, 매개 중심성, 근접 중심성, 고유벡터 중심성, 클러스터링 계수로 구성된다. 본 모형을 통해 일반 요인의 영향정도와 네트워크 구조요인의 영향정도를 파악하고 전체적으로 영향정도의 차이를 통합 분석하고자 한다.

〈Table 2〉 Conventional impact factors on view count of Youtube

Factor	Concept	Researchers
Comments	The number of users' comments on the content	Chatzopoulou et al. (2010), Borghol et al. (2012) Yoganarasimhan (2012)
Uploader Followers	The number of followers of the content uploader	Borghol et al. (2012)
Avg Rating	Average rating score by like(5) and dislike(1)	Chatzopoulou et al. (2010), Yoganarasimhan (2012), Susarla(2012)
Video age	The elapsed time after uploading the content	Borghol et al. (2012), Susarla(2012)

3.3 변수의 조작적 정의

기존 연구를 토대로 유튜브 조회 수 및 확산에 영향을 미치는 대표적인 일반요인은 <Table 2>과 같이 4개로 댓글 수(Comments), 등록자의 구독자 수(Uploader Followers), 평점(Avg Rating), 콘텐츠 경과시간(Video age)으로 정의하였다. 댓글 수(Comments)는 이용자들의 참여나 공감정도로 해석될 수 있다. 평점(Avg Rating)은 좋아요와 싫어요의 비율을 고려한 콘텐츠에 대한 평가 및 평균점수를 나타낸다. 평점은 유튜브에서 산정하는 방식으로 좋아요(5점)와 싫어요(1점)에 가중치를 부여하고 좋아요와 싫어요 빈도의 전체 합으로 나누어 0~5점으로 환산하여 계산된다. 그리고 등록자의 구독자 수(Uploader Followers)는 콘텐츠 등록자의 평판 및 인지도를 나타낸다고 볼 수 있다. 마지막으로 콘텐츠 경과시간(Video age)은 콘텐츠가 등록된 시점부터 얼마나 시간이 흘렀는지를 나타낸다.

추가적으로 조회 수와 네트워크 구조와의 관계를 살펴보기 위하여 <Table 3>과 같이 선행연구에서 언급된 5개의 네트워크 요인인 연결정도 중심성, 매개 중심성, 근접 중심성, 고유벡터 중심성, 클러스터링 계수를 사용하였다. 본 연구의 특성 상 무방향 네트워크를 사용하므로 연결정도 중심성의 내향연결정도(in-degree)와 외향연결정도(out-degree)는 산출하지 못하여 단순 연결정도 중심성만 사용하였다. 연결정도 중심성은 자신의 노드를 제외한 전체노드로 나누어 상대적인 연결정도 중심성으로 계산되었다.

〈Table 3〉 Network impact factors

Factor	Concept	Researchers
Degree Centrality	The number of links incident upon a node	Freeman(1979), Bolland(1988), Wasserman and Faust(1994), Costenbader and Valente(2003)
Closeness Centrality	The average of all distances of shortest paths between nodes	Freeman(1979), Bolland(1988), Wasserman and Faust(1994), Costenbader and Valente (2003)
Betweenness Centrality	The number of times a node acts as a bridge along the shortest path between two other nodes	Freeman(1979), Bolland(1988), Wasserman and Faust (1994), Costenbader and Valente (2003)
Eigenvector Centrality	a measure of the influence of a node in a network	Bonacich(1987), Bolland(1988), Wasserman and Faust(1994), Costenbader and Valente (2003)
Clustering Coefficient	a measure of the degree to which nodes in a graph tend to cluster together	WattsandStrogatz (1998), Girvan and Newman(2002)

4. 연구방법

4.1 연구대상 및 범위

본 연구에서는 콘텐츠 확산에 대한 영향 요인을 파악하는 것이기 때문에 활발하게 확산 및 검색되는 주제어를 모집단으로 선정할 필요가 있었다. 또한 일반화 가능성을 높이기 위하여 다양한 분야에서 균일하게 수집하여야 했다. 해당 조건을 만족하기 위하여 <Table 4>과 같이 구글의 카테고리별 최다 검색 주제어(2013년 10월 기준)를 키워드로 유튜브 콘텐츠를 수집하였다.

〈Table 4〉 Research scope for selecting contents

The most searched keywords in Google (Oct. 2013)				
Category	Sub Category	Top 1~5 ranked Keywords	Duplicated Keywords	Final Keywords
Business and Politics	7	35	17	218
Shopping	5	25		
Sports	8	40		
Entertainment	10	50		
Travel and Leisure	10	50		
Nature and Science	7	35		
Total	47	235	17	218

구글에서 제공하는 최다 검색어는 크게 6개로 구분되며 세부 카테고리는 총 47개이다. 세부 카테고리별 1~5위까지의 총 235개 주제어 중에서 중복되는 주제어 17개를 제외하였다. 따라서 최종적인 연구대상은 총 218개의 주제어를 기준으로 한 유튜브 콘텐츠로 확정되었다.

4.2 데이터 수집

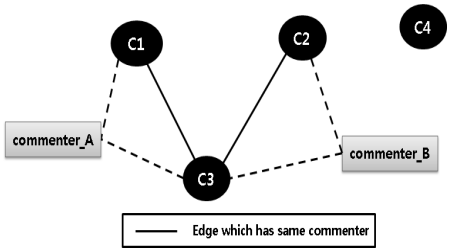
유튜브의 데이터 수집은 조회 수, 댓글 수 등과 같은 일반적인 데이터와 콘텐츠 중심의 무방향 네트워크 데이터로 구분하여 진행되었다. 그리고 특정 분야 및 카테고리에 대한 편향을 최소화하기 위하여 세부 카테고리별 100개의 동영상으로 제한하여 무작위로 수집하였다. 수집기간은 2013년 11월 27일부터 29일까지 총 3일 동안 진행되었다.

무방향 네트워크 데이터 구축 및 수집의 경우 콘텐츠 간 네트워크(video-video relatedness)를 사용하므로 유튜브의 검색 엔진을 기본으로 네트워크를 구성한다. 추가적으로 정확성을 보강하기 위하여 콘텐츠에 대해 동일한 댓글자가 있을 경우 유사 콘텐츠로 정의하고 네트워크를 연결하였다. 비슷한 방식은 과거 Santos et al.(2007) 등에 의해서 제시된 적이 있다. Santos et al.(2007)는 콘텐츠에 대한 정기 구독자 리스트를 이용하여 동일한 구독자가 있을 경우 유사 콘텐츠로 정의하고 네트워크를 연결하였다. 하지만 구독자 리스트보다 폭 넓고 사용이 용이한 댓글 데이터를 활용할 필요가 있었다.

최근에는 본 연구에서 사용한 방식과 동일한 네트워크 구축 방식을 Akrouf et al.(2013)가 제시 및 사용하였다. Akrouf et al.(2013)는 키워드를 통해 검색된 콘텐츠 안에서 동일 댓글자를 통한 네트워크를 구축하여 연구를 진행하였다. 이러한 네트워크 구축에 대한 논리는 다음과 같이 볼 수 있다. 첫째, 유튜브 검색엔진과 키워드를 통하여 실질적으로 유사한 콘텐츠들을 1차적으로 연결할 수 있다. 둘째, 특정 성향의 사람은 특정 콘텐츠를 이용하게 되므로 동일한 댓글자가 있는 콘텐츠들은 서로 유사할 가능성이 있다.

따라서 본 연구에서는 1차적으로 유튜브 검색엔진과 키워드를 이용하고 2차적으로 <Figure 2>과 같이 동일 댓글자를 기반으로 한 무방향 네트

워크를 구축하였다.



〈Figure 2〉 Content based Network Structuring

실질적인 데이터 수집은 1차적으로 MS Excel 기반 소셜 네트워크 분석 툴 NodeXL에 포함된 네트워크 데이터 수집 기능을 이용하였다. 이 기능을 통하여 동일 댓글자기반 콘텐츠 네트워크 데이터를 Edge List형태로 수집하였으며 조회수, 댓글 수 등과 같은 일반 데이터도 같이 수집하였다. 그리고 2차 데이터 수집은 등록자의 구독자 수, 좋아요, 싫어요와 같이 부족한 메타데이터를 보완하기 위하여 실시하였으며 프로그래밍 언어인 Python을 사용하여 자체적으로 제작하였다.

하지만 일부의 인기있는 콘텐츠에 한하여 실시간으로 빠르게 변하는 조회 수와 같은 데이터는 약간의 편차가 생기기 시작했다. 이를 개선하기 위하여 자체적으로 제작한 프로그램으로 2차 데이터 수집 시에 조회 수를 추가하여 재수집하였다. 또한 편차 및 오차를 최소화시키기 위하여 콘텐츠 주제어별로 1차 데이터 수집과 동시에 2차 데이터 수집을 실시하였다. 최종적으로 연구 데이터는 총 21,307개의 콘텐츠에 대한 일반 데이터와 네트워크 데이터로 수집되었다.

4.3 분석방법

데이터 분석은 '잔차분석', '다중공선성 진단', '회귀분석' 순으로 실시하여 시사점을 도출하는 단계로 진행되었다. 잔차분석은 히스토그램과 Q-Q도표를 통하여 정규성에 대해서 확인하였으며 다중공선성 진단은 상관분석과 VIF분석을 통하여 검증하였다. 또한 독립변수와 종속변수 간에 편상관분석을 통하여 회귀분석 결과에 대한 신뢰성 및 타당성을 확보한다. 회귀분석의 경우 전체 모집단에 대한 분석을 먼저 실시하고 확산속도 차이에 따라 인기 콘텐츠와 비인기 콘텐츠로 대상을 구분하여 추가적인 분석을 진행하였다. 네트워크 중심성 계산은 NodeXL을 이용하여 계산하였으며 실제 데이터의 특성을 고려하여 분석이 용이한 대표적인 통계 및 빅 데이터 분석 도구인 R을 이용하여 분석을 실시하였다.

5. 연구결과

5.1 기초통계량

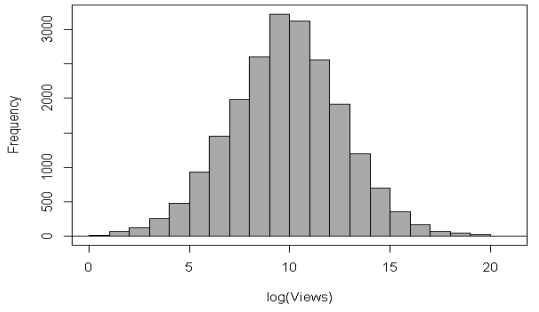
본 연구에 사용된 데이터의 기초통계량은 <Table 5>와 같다.

〈Table 5〉 Basic statistics

	N	Min	Max	Ave Rating	SD
Views	21307	0	937796879	968928	12116203
Comments	21307	0	6522165	1760	48082
Uploader Followers	21307	0	16911617	225967	886464
Video_age	21307	0	2857	346	432
Rating	21307	0.000	5.000	4.392	1.245
Degree Centrality	21307	0.000	0.263	0.017	0.030
Betweenness Centrality	21307	0.000	494124.061	520.144	7770.031
Closeness Centrality	21307	0.000	1.000	0.064	0.219
Eigenvector Centrality	21307	0.000	0.022	0.000	0.001
Clustering Coefficient	21307	0.000	1.000	0.187	0.302

5.2 정규성 검정

본 연구에 사용된 조회 수 등과 같은 데이터의 경우 특성 상 편차가 심하여 독립변수와 종속변수에 모두 log변환을 실시하였다. 따라서 정규성에 대해서 히스토그램과 Q-Q도표로 정규성을 나타내는지 재확인하였다. 인기도의 차이가 심한 유튜브의 특성 상 원래 데이터의 경우는 정규성을 만족하지 않았지만 log변환을 통한 데이터의 경우 아래 <Figure 3>과 같이 정규성을 만족하였다.



〈Figure 3〉 Histogram of log transformed view count

5.3 다중공선성 진단

다중공선성의 경우 일반적으로 상관분석을 통해 의심을 하며 VIF분석을 통하여 확정적인 진단을 하게 된다. 상관분석 결과 <Table 6>와 같이 대부분의 독립변수들은 서로 간에 높지 않은 상관관계를 가지는 것으로 나타났다. 하지만 실제로 분석에 사용된 log변환 데이터의 경우 <Table 6>와 같이 연결정도 중심성과 매개중심성이 0.66으로 약간 높게 나왔다. 하지만 다중공선성을 의심할 정도이며 다중공선성으로 판단하기에는 어렵다.

다중공선성 진단에서 대표적으로 사용되는 VIF분석의 경우 2.5를 넘으면 어느 정도 다중공선성 성향을 나타내며 10을 넘기면 거의 확정적인 것으로 판단한다. <Table 7>과 같이 분석결과 모든 독립변수의 VIF 값은 유효한 수준으로 다중공선성에 대하여 문제가 없는 것으로 나타났다. 특히, 상관분석에서 상관관계가 약간 높게 나타난 연결정도 중심성과 매개 중심성의 경우

〈Table 6〉 Correlation analysis on the log transformed independent variables

	Betweenness Centrality	Closeness Centrality	Clustering Coefficient	Comments	Degree Centrality	Eigenvecto Centrality	Uploader Followers	Ave Rating	Video Age
Betweenness Centrality	1	-0.18654	0.304143	0.035424	0.662519	0.046414	0.252834	0.029405	-0.11328
Closeness Centrality	-0.18654	1	-0.0906	-0.006	-0.1125	-0.01581	0.007309	0.00014	0.02568
Clustering Coefficient	0.304143	-0.0906	1	0.003025	0.501262	0.072611	0.148994	0.019653	-0.09228
Comments	0.035424	-0.006	0.003025	1	0.005617	-0.00198	0.064391	-0.03231	0.03045
Degree Centrality	0.662519	-0.1125	0.501262	0.005617	1	0.265872	0.225525	0.035016	-0.15842
Eigenvector Centrality	0.046414	-0.01581	0.072611	-0.00198	0.265872	1	0.040667	0.014457	-0.03965
Uploader Followers	0.252834	0.007309	0.148994	0.064391	0.225525	0.040667	1	0.057802	-0.18522
Rating	0.029405	0.00014	0.019653	-0.03231	0.035016	0.014457	0.057802	1	-0.00867
Video Age	-0.11328	0.02568	-0.09228	0.03045	-0.15842	-0.03965	-0.18522	-0.00867	1

VIF값이 모두 2.5를 넘지 않으므로 다중공선성에 대해 문제가 없는 것으로 판단할 수 있다.

〈Table 7〉 VIF analysis

Variable	VIF
Comments	1.033051
Uploader Followers	1.136724
Rating	1.004273
Video Age	1.060247
Degree Centrality	2.438050
Betweenness Centrality	1.940740
Closeness Centrality	1.042768
Eigenvector Centrality	1.120015
Clustering Coefficient	1.354174

5.4 편상관분석

다중공선성에 문제가 없더라도 독립변수 간 영향도 차이에 따라 회귀분석 결과가 다르게 나타날 수 있으므로 편상관분석(partial correlation)을 실시하여 결과에 대한 신뢰성과 타당성을 입증하고자 하였다.

편상관분석은 일반상관분석과는 다르게 다른 요인들의 영향력을 배제한 상태에서 측정된다. 이러한 상관분석을 통하여 다른 요인의 영향력을 포함하지 않았을 때 독립변수와 종속변수 간에 상관관계가 그 다음에 이루어질 회귀분석과 동일한 결과가 나오는지 확인하여 본 논문의 결과에 대한 신뢰성과 타당성을 보강한다. 편상관분석 결과, 대부분의 요인들은 조회 수와 정(+)의 관계를 갖는 것으로 나타났다. 하지만 연결정도 중심성과, 고유벡터 중심성은 <Table 8>과 같이 조회 수에 부(-)의 관계를 나타내는 것으로 분석되었다.

따라서 편상관분석에서 도출된 정(+)또는 부(-)의 결과가 동일하게 다음에 이루어질 회귀분석에서 도출된다면 독립변수 간에 영향도 및 관계 때문에 정(+) 또는 부(-)의 결과가 나온 것이 아님을 입증하며 타당함을 나타낸다.

〈Table 8〉 Partial correlation analysis on between independent variables and dependent variable

	Betweenness Centrality	Closeness Centrality	Clustering Coefficient	Comments	Degree Centrality	Eigenvector Centrality	Uploader Followers	Ave Rating	Video Age	Views
Betweenness Centrality	0	-0.17447	-0.0793	0.01166	0.60896	-0.17437	0.00314	-0.00175	-0.06999	0.24282
Closeness Centrality	-0.17447	0	-0.05428	-0.00956	0.03869	-0.01596	0.01105	0.00065	-0.01343	0.08241
Clustering Coefficient	-0.0793	-0.05428	0	-0.00382	0.41949	-0.08231	0.02254	-0.00007	-0.02114	0.0431
Comments	0.01166	-0.00956	-0.00382	0	-0.0177	0.00502	0.01751	-0.03801	0.01278	0.0811
Degree Centrality	0.60896	0.03869	0.41949	-0.0177	0	0.31592	0.04892	0.01252	-0.06438	-0.04394
Eigenvector Centrality	-0.17437	-0.01596	-0.08231	0.00502	0.31592	0	0.01525	0.00726	0.00546	-0.01403
Uploader Followers	0.00314	0.01105	0.02254	0.01751	0.04892	0.01525	0	0.03414	-0.29479	0.49837
Rating	-0.00175	0.00065	-0.00007	-0.03801	0.01252	0.00726	0.03414	0	-0.00163	0.02279
Video Age	-0.06999	-0.01343	-0.02114	0.01278	-0.06438	0.00546	-0.29479	-0.00163	0	0.33261
Views	0.24282	0.08241	0.0431	0.0811	-0.04394	-0.01403	0.49837	0.02279	0.33261	0

5.5 분석방법

데이터 분석은 '잔차분석', '다중공선성 진단', '회귀분석' 순으로 실시하여 시사점을 도출하는 단계로 진행되었다. 잔차분석은 히스토그램과 Q-Q도표를 통하여 정규성에 대해서 확인하였으며 다중공선성 진단은 상관분석과 VIF분석을 통하여 검증하였다. 또한 독립변수와 종속변수 간에 편상관분석을 통하여 회귀분석 결과에 대한 신뢰성 및 타당성을 확보한다. 회귀분석의 경우 전체 모집단에 대한 분석을 먼저 실시하고 확산 속도 차이에 따라 인기 콘텐츠와 비인기 콘텐츠로 대상을 구분하여 추가적인 분석을 진행하였다. 네트워크 중심성 계산은 NodeXL을 이용하여 계산하였으며 실제 데이터의 특성을 고려하여 분석이 용이한 대표적인 통계 및 빅 데이터 분석 도구인 R을 이용하여 분석을 실시하였다.

5.6 전체 모집단에 대한 회귀분석

회귀분석은 2차례로 나눠서 분석되었으며 먼저 수집된 전체 모집단에 대하여 진행되었다. 모집단 전체에 대한 회귀분석 결과는 <Table 9>와 같다.

전체 모집단에 대한 회귀분석 결과 기존에 알려진 일반요인들이 댓글 수, 등록자의 구독자 수, 평균 평점, 경과시간 모두 유의하고 조회 수에 정(+)의 영향을 미치는 것으로 나타났다. 영향력은 등록자의 구독자 수가 가장 높게 나타났으며 그 뒤로 경과시간, 댓글 수, 평균 평점 순으로 나타났다. 즉, 콘텐츠 등록자의 정기 구독자 수 또는 시간에 경과에 따른 인지도 등이 조회 수 에

〈Table 9〉 Regression result

Category	Non standard Estimate	Standard Beta	Std. Error	t value	Pr(>|t|)
log(Comments)	2.35E-05	0.14645536	9.06E-07	25.946	<2e-16 ***
log(Uploader Followers)	3.34E-01	0.45914615	4.31E-03	77.546	<2e-16 ***
log(Rating)	4.15E-01	0.01865470	1.24E-01	3.352	0.000804 ***
log(Video_Age)	4.83E-01	0.27436280	1.01E-02	47.979	<2e-16 ***
log(Degree. Centrality)	-4.19E+00	-0.04790651	7.58E-01	-5.525	3.34E-08 ***
log(Betweenness. Centrality)	2.74E-01	0.26743918	7.92E-03	34.568	<2e-16 ***
log(Closeness. Centrality)	1.08E+00	0.06745270	9.09E-02	11.894	<2e-16 ***
log(Eigenvector. Centrality)	-3.48E+01	-0.01195665	1.71E+01	-2.034	0.041927 *
log(Clustering. Coefficient	4.31E-01	0.03792200	7.34E-02	5.868	4.48E-09 ***
Multiple R-squared: 0.3963 Adjusted R-squared: 0.3961 F-statistic: 1428 ***					
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

가장 큰 영향을 미치는 것으로 분석되었다. 그리고 댓글 수도 조회 수에 긍정적인 영향을 미치나 평점은 상대적으로 적은 영향을 미치는 것으로 분석되었다.

네트워크 요인들의 경우도 영향력을 나타내는 고유벡터 중심성이 뚜렷하게 나타나지는 않았지만 모든 요인들이 유의한 것으로 나타났다. 연결정도 중심성과 고유벡터 중심성의 경우 앞서 실시한 편상관분석과 동일하게 조회 수에 부(-)의 영향을 미치는 것으로 나타났다. 이러한 결과는 연결정도 중심성이 높은 경우 유사 콘텐츠들이 많이 존재하며 해당 콘텐츠를 조회할 필요성이 없어지고 조회 수가 분산되는 것으로 분석된다. 또한 고유벡터 중심성이 높은 경우는 영향력 있는 인기 콘텐츠가 연결되어 있어 해당 콘텐츠에 조회 수를 빼앗기는 것으로 분석된다.

매개 중심성과 근접 중심성의 경우 정(+)의 영향을 미치는 것으로 나타났다. 매개 중심성이 높은 경우 특정 콘텐츠를 보고 해당 콘텐츠에 등장하는 대상이나 유사 콘텐츠에 대한 검색을 유발하는 것으로 분석된다. 근접 중심성이 높은 경우는 매개중심성과 유사하지만 매개역할이 아닌 네트워크 구조적인 거리를 줄여주게 되므로 조회 수가 느는 것으로 분석된다. 즉, 일정한 양의 정보를 얻기 위하여 다수의 콘텐츠를 검색 시 검색빈도를 줄여주는 콘텐츠가 조회 수 확보에 유리한 것으로 분석된다.

마지막 클러스터링 계수의 경우도 정(+)의 영향을 미치는 것으로 나타났다. 클러스터링 계수는 밀집도를 나타내며 연결된 콘텐츠들 간에 긴밀하게 연결되어 있는 경우에 나타난다. 즉, 독단적인 콘텐츠가 아니라 특정 주제로 군집되어 있는 콘텐츠의 경우 조회 수에 긍정적인 시너지 효과를 가지고 오는 것으로 분석된다.

이와 같이 회귀분석 결과 전체적으로는 등록자의 구독자 수, 경과시간, 매개 중심성, 댓글 수, 근접 중심성, 클러스터링 계수, 평균 평점 순으로 대부분의 요인들이 조회 수에 긍정적인 영향을 미치지만 연결정도 중심성과 고유벡터 중심성은 부정적 영향을 주는 것으로 분석되었다.

6. 결 론

유튜브는 기존 폐쇄형 콘텐츠 서비스와는 다르게 개방형 콘텐츠 서비스로 이용자들의 참여와 공유를 통하여 많은 인기를 유지하고 있다. 하지만 이러한 오픈 콘텐츠 서비스 형태가 구조적인 측면에서 조회 수에 긍정적인 영향뿐만 아니라 부정적 영향도 줄 수 있을 것이다. 본 연구는 커다란 인기를 가지고 있는 유튜브 콘텐츠의 효과적인 활용을 위해서 무엇이 필요한지에 대한 고민 속에서 진행되었다.

기존의 연구는 일반적인 요인과 네트워크구조 요인을 별개로 영향요인을 분석하였다. 하지만 본 논문에서는 일반요인과 네트워크 구조 요인을 통합하여 조회수의 영향요인을 분석하였다. 또한, 네트워크 구조를 사람중심의 네트워크가 아니라 댓글기반으로 유투브상의 21,307개의 실질적인 콘텐츠 중심의 네트워크를 구성하여 분석하였다.

본 연구를 통해 기존에 알려진 일반적 요인과 네트워크 요인들이 모두 조회수에 영향을 주는 인과관계를 통계적으로 재검증하였으며 통합적으로는 등록자의 구독자 수, 경과시간, 매개 중심성, 댓글 수, 근접 중심성, 클러스터링 계수, 평균 평점 순으로 조회 수에 긍정적인 영향을 미치는 것으로 분석되었다. 하지만 네트워크 요인중

연결정도 중심성과 고유벡터 중심성은 부정적 영향을 주는 것으로 분석되었다.

이를 통해서 콘텐츠 확산에 활용할 전략적 시사점은 다음과 같다. 첫째, 등록자의 구독자 수, 경과시간, 댓글 수, 평균 평점 등 일반요인의 관리 모두 필요하다. 하지만 평점평균은 상대적으로 크게 신경쓰지 않아도 된다. 이와 반면에 등록자의 구독자 수를 늘리는 전략이 필요하다. 또한 콘텐츠의 등록 시간을 오래 유지할수록 유리하다.

둘째, 네트워크 요인들중 매개 중심성과 근접 중심성의 영향력에 주의해야 한다. 특정 콘텐츠를 보고 해당 콘텐츠에 등장하는 대상이나 유사 콘텐츠에 대한 검색을 유발하는 것으로 분석된다. 다른 인기 콘텐츠와의 거리를 줄여주는 것이 필요하다. 즉, 콘텐츠 검색 시 검색빈도를 줄여주고 다수의 콘텐츠와 유사도를 높이는 전략이 조회 수 확보에 유리한 것으로 분석된다. 이것은 클러스터 계수의 영향분석 결과와도 일치하는 것이다.

셋째, 하지만 네트워크 요인중 연결정도 중심성과 고유벡터 중심성이 부(-)의 영향을 준다는 점에 유의해야 한다. 즉, 너무 많은 콘텐츠와 연결성을 높일 경우 유사 콘텐츠들이 많이 존재하여 조회 수가 분산될 수 있으므로 주의해야 한다. 또한 고유벡터 중심성이 높은 경우는 주위에 영향력 있는 인기 콘텐츠가 연결되어 있어 해당 콘텐츠에 조회 수를 빼앗길 수 있으므로 너무 인기 있는 콘텐츠에 연결되는 것은 불리할 수 있다.

본 연구를 통하여 유튜브 콘텐츠 확산에 대한 일반적인 영향요인과 구조적인 현상을 통합적으로 규명하였다. 사회적인 측면에서는 음반산업의 수요예측이나 콘텐츠 제작 업체들의 원활한 서비스 제공을 위한 설명력있는 영향요인 및 모델을 제시할 수 있을 것이다. 또한 기업들이 유튜브와 같은 콘텐츠 서비스를 통한 온라인 마케팅 활동 시 콘텐츠 제공의 전략적 근거를 제공하였다.

본 연구의 한계점으로는 키워드를 기반으로 한 동일 댓글자 네트워크로 콘텐츠 구조를 설명하였지만 이러한 네트워크 구축 방식은 간접적인 콘텐츠 네트워크 구조를 볼 수는 있다. 향후 좀 더 다양한 방식의 직접적인 콘텐츠 네트워크 구축방식도 활용할 필요가 있다. 향후 연구주제로는 콘텐츠의 형태나 성격별로 영향요인을 분석하는 등 관련된 다양한 연구가 가능할 것이다.

참 고 문 헌 (References)

Akrouf, S., L. Meriem, B. Yahia, and M. N. Eddine, “Social Network Analysis and Information Propagation: A Case Study Using Flickr and YouTube Networks,” *International Journal of Future Computer and Communication,* Vol. 2, No. 3(2013), 21~22.

Bolland, J. M., “Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks,” *Social Networks*, Vol. 10, No. 3(1988), 233~253.

Bonacich, P., “Power and centrality: A family of measures,” *American Journal of Sociology*, Vol. 92, No. 5(1987), 1170~1182.

Borghol, Y., S. Ardon, N. Carlsson, D. Eager, and A. Mahanti, “The untold story of the clones: content-agnostic factors that impact youtube video popularity,” *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining,*

(2012), 1186~1194.

Brin, S. and L. Page, "The anatomy of a large-cale hypertextual Web search engine," *Computer networks and ISDN systems,* Vol. 30, No. 1(1998), 107~117.

Burgess, J. E. and J. B. Green, *YouTube: Online video and participatory culture*, Cambridge: Polity press, 2009.

Chatzopoulou, G., C. Sheng, and M. Faloutsos, "A first step towards understanding popularity in youtube," *INFOCOM IEEE Conference on Computer Communications Workshops,* (2010), 1~6.

Cho I. D and N. G. Kim, "Recommending Core and Connecting Keywords of Research Area Using Social Network and Data Mining Techniques," *Journal of Intelligence and Information Systems,* Vol.17 No.1(2011), 127~138.

Costenbader, E. and T. W. Valente, "The stability of centrality measures when networks are sampled," *Social networks,* Vol. 25, No. 4(2003), 283~307.

Freeman, L. C., "Centrality in social networks conceptual clarification," *Social networks,* Vol. 1, No. 3(1979), 215~239.

Girvan, M. and M. E. Newman, "Community structure in social and biological networks," *roceedings of the National Academy of Sciences*, Vol. 99, No. 12(2002), 7821~7826.

Kaplan, A. M. and M. Haenlein, "ers of the world, unite! The challenges and opportunities of Social Media," *Business horizons,* Vol. 53, No. 1(2010), 59~68.

Kim, I. J., D. C. Lee, and G. G. Lim, "A study on the Critical Success Factors of social Commerce through the Analysis of the Perception Gap between the Service Providers and the Users: Focused on Ticket Monster in Korea", *Asia Pacific Journal of Information Systems,* Vol. 24, No. 2(2014), 211~232.

Lee, S., S. Park, G, G. Lim, and S. Baek, "A Roadmap for Developing Digital Content Distribution Infrastructure," *Journal of Korea Society of IT Services,* Vol. 8, No. 4(2009), 75~86.

Park B. E. and G. G. Lim, "A Study on the Diffusion Impact Factors through Contents Network Analysis : Focused on Youtube," *Proceedings of the Korea Inteligent Information System Society Conference*, (2015).

Pallis, G., D. Zeinalipour-Yazti, and M. D. Dikaiakos, *New Directions in Web Data Management 1*, Springer Berlin Heidelberg. (2011), 213~234.

Santos, R. L., B. P. Rocha, C. G. Rezende, and A. A. Loureiro, "Characterizing the YouTube video-sharing community," *Technical Report,* Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil, 2007.

Susarla, A., J. H. Oh, and Y. Tan, "Social networks and the diffusion of user-generated content: Evidence from YouTube," *Information Systems Research,* Vol. 23, No. 1(2012), 23~41.

Wasserman, S. and K. Faust, *Social network analysis: Methods and applications*, Cambridge university press, Vol. 8, 1994.

Watts, D. J. and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, Vol. 393, No. 6684(1998), 440~442.

Yoganarasimhan, H., "Impact of social network structure on content propagation: A study using YouTube data," *Quantitative Marketing and Economics,* Vol. 10, No. 1(2012), 111~150.

Abstract

A Study on the Impact Factors of Contents Diffusion in Youtube using Integrated Content Network Analysis

Byung Eun Park* · Gyoo Gun Lim**

Social media is an emerging issue in content services and in current business environment. YouTube is the most representative social media service in the world. YouTube is different from other conventional content services in its open user participation and contents creation methods. To promote a content in YouTube, it is important to understand the diffusion phenomena of contents and the network structural characteristics. Most previous studies analyzed impact factors of contents diffusion from the view point of general behavioral factors. Currently some researchers use network structure factors. However, these two approaches have been used separately. However this study tries to analyze the general impact factors on the view count and content based network structures all together. In addition, when building a content based network, this study forms the network structure by analyzing user comments on 22,370 contents of YouTube not based on the individual user based network.

From this study, we re-proved statistically the causal relations between view count and not only general factors but also network factors. Moreover by analyzing this integrated research model, we found that these factors affect the view count of YouTube according to the following order; Uploader Followers, Video Age, Betweenness Centrality, Comments, Closeness Centrality, Clustering Coefficient and Rating. However Degree Centrality and Eigenvector Centrality affect the view count negatively.

From this research some strategic points for the utilizing of contents diffusion are as followings. First, it is needed to manage general factors such as the number of uploader followers or subscribers, the video age, the number of comments, average rating points, and etc. The impact of average rating points is not so much important as we thought before. However, it is needed to increase the number of uploader followers strategically and sustain the contents in the service as long as possible.

* Business School, Hanyang University
** Corresponding Author: Gyoo Gun Lim
Business School, Hanyang University
222 Wangsimni-ro, SeongDong-gu,Seoul,133-791, Korea
Tel: +82-2-2220-2593, Fax: +82-2-2220-1169, E-mail: gglim@hangyang.ac.kr

Second, we need to pay attention to the impacts of betweenness centrality and closeness centrality among other network factors. Users seems to search the related subject or similar contents after watching a content. It is needed to shorten the distance between other popular contents in the service. Namely, this study showed that it is beneficial for increasing view counts by decreasing the number of search attempts and increasing similarity with many other contents. This is consistent with the result of the clustering coefficient impact analysis.

Third, it is important to notice the negative impact of degree centrality and eigenvector centrality on the view count. If the number of connections with other contents is too much increased it means there are many similar contents and eventually it might distribute the view counts. Moreover, too high eigenvector centrality means that there are connections with popular contents around the content, and it might lose the view count because of the impact of the popular contents. It would be better to avoid connections with too powerful popular contents.

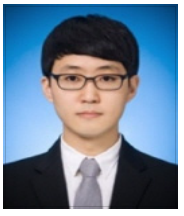
From this study we analyzed the phenomenon and verified diffusion factors of Youtube contents by using an integrated model consisting of general factors and network structure factors. From the viewpoints of social contribution, this study might provide useful information to music or movie industry or other contents vendors for their effective contents services. This research provides basic schemes that can be applied strategically in online contents marketing.

One of the limitations of this study is that this study formed a contents based network for the network structure analysis. It might be an indirect method to see the content network structure. We can use more various methods to establish direct content network. Further researches include more detailed researches like an analysis according to the types of contents or domains or characteristics of the contents or users, and etc.

Key Words : Content Network; Youtube; Difussion Factors; Social Media; Influencing Factor

Received : June 11, 2015 Revised : September 7, 2015 Accepted : September 9, 2015
Type of Submission : Normal Track Corresponding Author : Gyoo Gun Lim

저 자 소 개



박 병 연
한양대학교 대학원 경영학과에서 경영정보시스템 전공으로 석사학위를 취득하였다. 주요 연구분야는 머신러닝, 빅데이터 플랫폼, 타겟 마케팅이며, 특히 데이터 기반 플랫폼 활성화에 관심을 두고 있다. 최근 (주)에코마케팅에서 데이터마이닝을 연구하고 있다.

임 규 건
KAIST 전산학 학사, POSTECH 전자계산학 석사, KAIST 경영공학 박사학위를 취득하였고, 삼성전자, KT 연구개발본부 전임연구원, 국제전자상거래연구센터(ICEC)의 연구위원, 세종대학교 경영학과 부교수를 거쳐 현재 한양대학교 경영대학 교수로 재직하고 있다. 한국지능정보시스템학회 부회장, 한국IT서비스학회 편집위원장, 한국전자거래학회 이사, UCI 운영위원 등의 활동을 하고 있다. 주요 저서로는 경영을 위한 정보기술(2007, 교보문고), e-비즈니스 경영(2005, 이프레스), 디지털경제시대의 경영정보시스템(2003, 사이텍미디어) 등이 있으며, 관심분야는 혁신비즈니스모델, IT 서비스 경영, e-Business, MIS, Intelligent Systems 등 이며 다수의 프로젝트 참여 경력과 논문과 관련 특허가 있다.