

Technische Universität Berlin  
Seminar: Internet Measurement  
Betreuer: Fabian Schneider

YouTube Traffic Characterization: A View From the  
Edge“  
/  
„I Tube, You Tube, Everybody Tubes: Analyzing the  
World's Largest User Generated Content Video  
System

Zusammenfassung:

Diese Arbeit fasst die beiden Papers „YouTube Traffic Characterization“[1] und „I Tube, You Tube, Everybody Tubes“[2] zusammen. In den letzten Jahren hat sich das Internet von statischen Webseiten, auf Webseiten mit interaktivem Charakter erweitert. Dieses ermöglicht dass eine große Nutzeranzahl in der Lage ist Inhalte zu publizieren und zu bewerten. Diese neue Art von Web-Traffic wird im Folgenden auf vorhandene Charakteristiken der übertragenen Dateien, Popularität der angeforderten Dateien und Aufrufhäufigkeit der Dateien untersucht. Es werden Unterschiede zu Web 1.0-Traffic genannt und das Potential, dass die Web 2.0-Charakteristiken für Caching bieten, untersucht.

Irina Antonova  
(irinaa@cs.tu-berlin.de)

WS 2008/2009 (23. Februar 2009)

## 1. Einleitung

Das Internet hat sich in den letzten Jahren grundlegend verändert. In den Anfängen haben Nutzer hauptsächlich auf statische Webinhalte zugegriffen. Diese Webinhalte wurden von Unternehmen, Universitäten oder anderen Organisationen im Internet publiziert. Der durchschnittliche Nutzer verfügte nicht über das notwendige Wissen um Inhalte ins Netz zu stellen. Diese Art von Web-Inhalten wird im Weiteren als Web1.0 bezeichnet.

Das jetzige Netz beinhaltet zusätzlich Websites, die es auch technisch nicht versierten Nutzern ermöglichen, einfach eigene Inhalte, von Texten bis Videos, ins Internet zu stellen. Beispiele hierfür sind Seiten wie Flickr<sup>1</sup>, Facebook<sup>2</sup> oder YouTube<sup>3</sup>, auf denen Nutzer an sozialen Netzwerken teilnehmen und eigene Beiträge publizieren können. Diese Art von Webseiten, im Weiteren als Web2.0 bezeichnet, bietet weitere Interaktionsmöglichkeiten wie Tagging und Rating.

*Tagging* ist eine Funktion die es Nutzern erlaubt von ihnen publizierte Inhalte mit Worten oder Phrasen zu verknüpfen, so dass eine Strukturierung anhand dieser Begriffe erfolgt.

*Rating* gibt den Nutzern die Möglichkeit Inhalte anderer Nutzer innerhalb einer Bewertungsskala einzuordnen.

Das Web2.0 hat im Gegensatz zum Web1.0 also eine geringere Eintrittsbarriere. Nutzer aller Altersgruppen können Inhalte aller Art einfach publizieren, bewerten und kommentieren. Diese Entwicklung, zusammen mit der vermehrten Verbreitung von Breitbandzugängen in privaten Haushalten, hat zur Folge dass das Datenvolumen im Netz rapide ansteigt und Netzwerkressourcen stark ausgelastet werden. Um effiziente Netzwerk-Management- und Kapazitätsplanungsstrategien zur optimalen Ausnutzung von Bandbreite und Netzwerkressourcen entwickeln zu können, müssen zunächst die Charakteristiken des Web2.0-Traffic untersucht werden.

Hierzu wurde ein Teil des Web-Traffics der Webseite YouTube genauer betrachtet. YouTube wurde als Untersuchungsgegenstand gewählt, da es die größte Video-Sharing Website ist[3], über welche etwa 60 Prozent aller im Netz geschauten Videos geladen werden. Mit 65 000 Video Uploads pro Tag ist sie auch eine der am schnellsten wachsenden Seiten.

## 2. Methodologie

Da die auf YouTube verfügbaren Web-Inhalte nicht statisch sind, ist es für eine Charakterisierung notwendig Web-Traffic über eine bestimmte Zeitspanne zu beobachten um brauchbare Aussagen zu erlangen. Ein weiteres Problem ist, dass die zu beobachtenden Datenmengen und der dazugehörige Daten-Traffic in einer Größenordnung sind, die es unmöglich macht YouTube-Verhalten global aufzuzeichnen. Aus diesen beiden Gründen wurde sich im Paper „You Tube Traffic Characterization“[1] für eine 3-monatige Betrachtung von YouTube-Verhalten entschieden. In dieser wird aus der lokalen Perspektive in einem Universitätsnetzwerk der gesamte YouTube-Traffic

---

<sup>1</sup> [www.Flickr.com](http://www.Flickr.com)

<sup>2</sup> [www.facebook.com](http://www.facebook.com)

<sup>3</sup> [www.youtube.com](http://www.youtube.com)

aufgezeichnet und genauer betrachtet. Und zum anderen werden aus globaler Perspektive die Charakteristiken der „Most Popular“ Videos auf der YouTube-Website selbst untersucht.

### 2.1 Campus Netzwerk(Lokale Perspektive)

Auf dem betrachteten Campus wohnen und arbeiten 33 300 Personen, 28 000 davon sind Studenten, der Rest Lehrpersonal und Angestellte.

Die relevanten Web-Traces wurden folgenderweise protokolliert:

1. Es wurden alle Server-IPs von Servern, von denen YouTube Inhalte geladen wurden, herausgefiltert.
2. Mit Hilfe von bro<sup>4</sup> werden alle Transaktionen, die die aus 1. identifizierten IPs enthalten, protokolliert. Wobei in Echtzeit eine Zusammenfassung der Transaktion extrahiert wird.
3. Aufgrund der großen Datenmenge, muss bro jeden Tag neu gestartet werden. Die Log-Datei wird komprimiert und abgelegt.

### 2.2 YouTube(Globale Perspektive)

Um Informationen über Datei-Charakteristiken zu erhalten, werden über den 3-monatigen Untersuchungszeitraum täglich Informationen über die „Top 100“ der meistaufgerufenen YouTube Beiträge pro Tag, Woche, Monat und Insgesamt, von der YouTube-Website heruntergeladen. Hierfür wird eine von YouTube für Entwickler bereitgestellte API<sup>5</sup> genutzt, mit welcher anhand der auf den Top 100 Seiten gefundenen Video-IDs, zusätzliche Statistiken zum Video erhoben werden können. Diese Videos bilden nur einen Anteil der auf YouTube verfügbaren Videos. Nach dem Pareto Prinzip[4] jedoch, verursachen auf Web- und Medien-Servern 10 Prozent (bzw. 20 Prozent) der verfügbaren Inhalte 90 Prozent (bzw. 80 Prozent) der Anfragen.

Im Paper „I Tube, You Tube, Everybody Tubes“ [ 2] hingegen, wurden Eigenschaften aller Videos aus der Kategorie „Science & Technology“ untersucht. Hierfür wurden innerhalb eines 6-tägigen Untersuchungszeitraums Eigenschaften der Videos wie unter anderem Alter, View-Anzahl und Rating durch einen Crawler an jedem der Sechs Tage protokolliert. In dieser Arbeit liegt der Fokus mehr in der Popularitätsentwicklung und – Verteilung der Videos.

## 3. Auswertung

Im Folgenden werden die im Datenbestand identifizierten Charakteristiken von Web2.0 Traffic vorgestellt.

---

<sup>4</sup> Bro ist ein open-source, Unix-based Network Intrusion Detection System (NIDS) das Netzwerk Traffic protokollieren und filtern kann.

<sup>5</sup> <http://www.youtube.com/dev>

### 3.1 Datenbestand(Paper [ 1])

Die Datenaufnahme dauerte 85 Tage, vom 14.Jan.2007 bis zum 8.Apr.2007. Dieser Zeitraum beinhaltet verschiedene Zeiträume, wie z.B. den Semesteranfang, den Prüfungszeitraum und die Vorlesungsfreie Zeit, so dass die beobachteten Charakteristiken nicht zu stark von Eigenheiten des betrachteten Zeitraums abhängig sind. Es wurden 23 250 438 Transaktionen protokolliert und 6,54 TB an Daten übertragen. Es gab insgesamt 625 593 Videoanfragen mit einem Datenvolumen von 6,45 TB, von diesen waren 323 677 unterschiedliche Videoanfragen.

Aus Tabelle 1 ist ersichtlich, dass Videoanfragen und –antworten 98,9 Prozent des übertragenen Datenvolumens ausmachen. Nur etwa die Hälfte aller angefragten Videos(bzw. übertragener Bytes) sind unterschiedlich. Mit effektiven Caching-Strategien könnte der YouTube bezogene Web-Traffic um annähernd die Hälfte reduziert werden.

Tabelle 1: Zusammenfassung des lokalen YouTube Datenbestands

| Item                     | Information   |
|--------------------------|---------------|
| Start Date               | Jan. 14, 2007 |
| End Date                 | Apr. 8, 2007  |
| Total Valid Transactions | 23,250,438    |
| Total Bytes              | 6.54 TB       |
| Total Video Requests     | 625,593       |
| Total Video Bytes        | 6.45 TB       |
| Unique Video Requests    | 323,677       |
| Unique Video Bytes       | 3.26 TB       |

In Tabelle 2 sieht man dass nahezu 100% der HTTP-Anfragen GET Anfragen sind, nur 0,12% sind POST Anfragen. Der größte Teil der übertragenen Daten wird also von Konsumenten verursacht.

Tabelle 2: Verteilung der HTTP-Anfragetypen

| Method | Total      | % of Total |
|--------|------------|------------|
| GET    | 23,221,168 | 99.87      |
| POST   | 28,655     | 0.12       |
| Others | 615        | 0.01       |

### 3.2 Video Charakteristiken

Im Folgenden werden die Videoeigenschaften: Dateigröße, Videodauer, Bitrate und Alter der Videos betrachtet.

#### 3.2.1 Videodateigrößen

Die Videodateigrößen im Vergleich zur Größe anderer Dateien wurde anhand des Content-Length-Feldes der HTTP Header bestimmt.

Festzustellen ist, dass 2,8 Millionen(bzw. 0,3 Millionen) verschiedene Bilder- (bzw. Videodateien) übertragen wurden, jedoch insgesamt weniger als 1200 Text- und Anwendungsdateien. Es wird ein Standard-Framework von HTML und Javascript-Seiten zur Darstellung der Videos genutzt. Die hohe Bilderanzahl erklärt sich durch die Nutzung von verschiedenen Icons auf den YouTube-Seiten.

Grundsätzlich begrenzt YouTube die Größe der hochladbaren Videos auf eine Länge von höchstens 10 Minuten bzw. eine Dateigröße von nicht mehr als 1GB.<sup>6</sup>

Wie in Abbildung 1 zu erkennen liegt die typische Dateigröße von Videos zwischen 500KB und 100000KB, also deutlich höher als die Dateigrößen anderer Dateitypen, wie Text, Bilder oder Anwendungen.

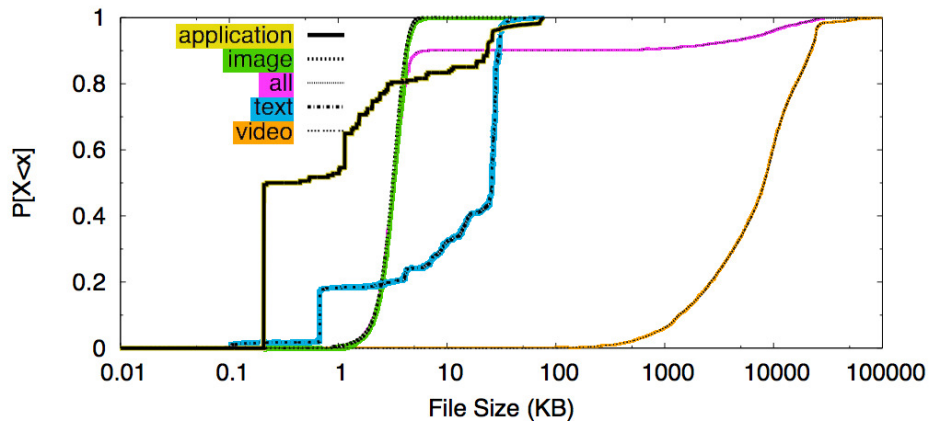


Abbildung 1: CDF eindeutiger Dateigrößen

### 3.2.2 Videodauer

Die Videodauer der angefragten Videos wurde durch eine von YouTube bereitgestellte API erfragt, da die HTTP-Header diese Information nicht enthalten. Zunächst ist festzustellen, dass trotz der YouTube vorgegebenen 10-Minutengrenze, doch eine geringe Anzahl längerer Videos vorhanden sind. Da 99,9% der Videos eine geringere Länge als 2 Stunden aufweisen wurde der Untersuchungsbereich aus [1] auf diese Videos eingeschränkt. Die durchschnittliche Videodauer der angefragten Videos liegt bei 4,15 Minuten, der gewichtete Durchschnitt bei 3,33 Minuten. Videos die langfristig in „Most Popular“-Listen auftauchen habe grundsätzlich eine etwas kürzere Dauer.

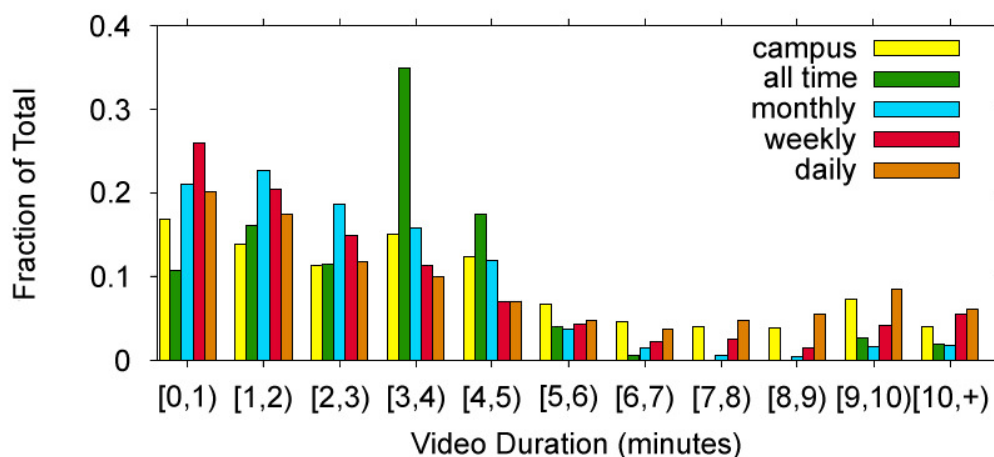


Abbildung 2: Histogramm der Dauer der Videos

<sup>6</sup> <http://www.google.com/support/youtube/bin/answer.py?answer=55743&%20topic=10527>

In Abbildung 2 ist zu sehen, dass der Anteil der Videos in den täglichen „Most Popular“-Listen am Größten ist bei einer Videolänge von 0 bis zu 1 Minute. Je länger die Videodauer, desto geringer der Anteil der Videos an täglichen „Most Popular“-Listen. Bei den „All Time Most Popular“-Listen ist die die am häufigsten anzutreffende Videolänge zwischen 3 und 4 Minuten. Diese Videos sind die am häufigsten aufgerufen und liegen somit nah um den ermittelten gewichteten Durchschnitt von 3,33 Minuten. [Hinweis zu Abbildung 2: Reihenfolge der Kategorien „Campus“, „All Time“, ... entspricht in jedem Intervall, der Anordnung in der Legende]

### 3.2.3 Bitrate

Die Bitrate der Videos wurde anhand der bereits vorhandenen Videodateigröße und der Videodauer geschätzt. Sie ist relevant, da bei einer Vergrößerung der für Nutzer verfügbaren Bandbreite, vermehrt Videos mit besserer Qualität, d.h. größerer Bitrate angefragt werden würden. Dieses würde die Dateigröße der Videos erhöhen. Die Untersuchung am Campus ergab, dass nur wenige Videos mit geringen Bitraten codiert waren, genauso wie nur wenige Videos Bitraten von mehr als 1Mbps hatten. 97% der geladenen Videos hatten Bitraten kleiner als 1Mbps. Da 62,6% der Videos Bitraten zwischen 300 und 400Kbps hatten liegt es nahe, dass YouTube als Zielgruppe Breitbandnutzer hat, die bei diesen Bitraten ohne Verzögerung die angeforderten Videos abspielen können.

### 3.2.3 Alter der Videos

Das Alter der angefragten Videos liegt bei 5% der Anfragen bei mehr als einem Jahr. 73% der angefragten Videos waren älter als einen Monat. Daraus folgt der Schluss, dass Nutzer sich durchaus auch für ältere Inhalte interessieren. Des Weiteren wird betrachtet welche der angeforderten Dateien kürzlich geupdated („recently modified“) wurden. Ergebnis war, dass Video- und Bilderdateien sich im Gegensatz zu Text- und Anwendungsdateien relativ statisch verhalten.

## 3.3 Popularität der Videos

Die Popularität von Dateien steht meist in direktem Zusammenhang mit der Aufrufhäufigkeit dieser Dateien. Um den Zusammenhang zwischen der Popularität und Aufrufhäufigkeit bei YouTube-Videos zu untersuchen werden 2 verschiedene Verfahren betrachtet. Die Zipf Analyse und die Konzentrationsanalyse.

### 3.3.1 Zipf Regel

Die Zipf Regel besagt, dass es einen Zusammenhang zwischen der Vorkommenshäufigkeit von Objekten und deren Popularität / deren Rang existiert.

Die Anwendbarkeit der Zipf-Regel wird folgendermaßen überprüft:

- es werden alle Objekte gemäß ihrer Popularität angeordnet und dem populärsten Objekt Rang 1(R) vergeben, dem zweit-populärsten Objekt wird Rang 2(R) vergeben usw.

- die Vorkommenshäufigkeit(F) des Objektes entspricht dann:

$$F \sim R^{-\beta}$$

- $\beta$  ist eine nah an 1 liegende Konstante

Plottet man nun die Vorkommenshäufigkeit F und den Rang R der Objekte in ein Koordinatensystem mit logarithmischen Skalen, so ist das auftreten einer geraden Linie ein Indiz für das Erfüllen der Zipf Regel. Die untenstehende Grafik bildet Vorkommenshäufigkeit und Rang der am Campus untersuchten Videoobjekte ab und zeigt eine Zipf-ähnliche Verteilung. Der Rang eines Videos wurde festgelegt anhand der Anzahl der Requests mit der eindeutigen ID dieses Videos. Die ist möglich, da YouTube das Runterladen von Videos nicht unterstützt, und somit für jedes Mal abspielen eines Videos, ein neues Get-Request erfolgen muss. Trifft die Zipf Regel zu bietet es sich an Video-Caching-Strategien zu entwickeln[5].

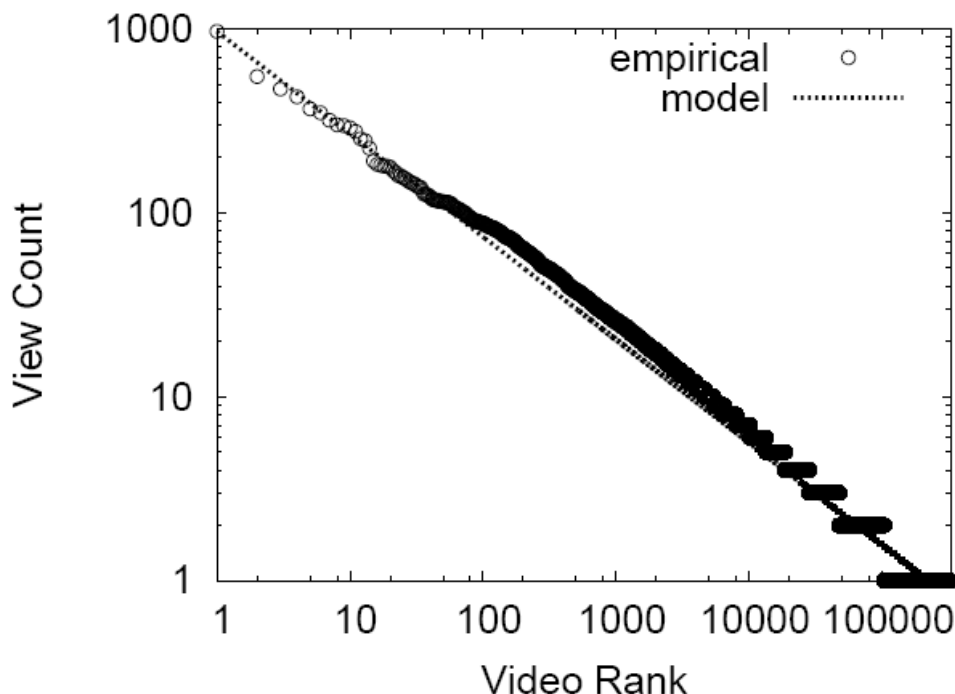


Abbildung 3: Bewertete Aufrufanzahl der Videos (Campus)

### 3.3.2 Konzentrationsanalyse

Die Konzentrationsanalyse dient dazu zu verstehen, wie groß der Anteil von Referenzen, die auf „Most Popular“-Videos zurückgehen, an der Gesamtmenge der vorhandenen Referenzen auf dem Campus ist.

In den Campus-Traces verursachen die meistaufgerufenen 10% (bzw. 20%) der Videos jedoch nur 40% (bzw. 52%) der Videoaufrufe. Offensichtlich gilt die Pareto-Regel[4] in diesem Fall nicht. Das lässt sich durch die Fülle und Diversität der bei YouTube vorhandenen Videos erklären, die teilweise sehr spezifische Inhalte haben und somit nur für bestimmte Zielgruppen interessant sind. Die Menge der vorhandenen Videos führt dazu dass eine geringere Konzentration der Anfragen auf die Videos der „Most Popular“-

Listen vorhanden ist. Allerdings ist es auch möglich, dass die Campusbewohner hauptsächlich von Freunden empfohlene Videos bzw. verstärkt Nischen-Videos mit Bezug zur Universität sehen. Derartige, lokale, Videos tauchen in den globalen „Most Popular“-Listen natürlich nicht auf. Hierfür spricht auch, dass im Paper[2] Teilkategorien von YouTube mit einem Crawler untersucht wurden, mit dem Ergebnis, dass 10% der beliebtesten Videos 80% der Aufrufe verursachen. Global gesehen könnte die Pareto-Regel also durchaus zutreffend sein.

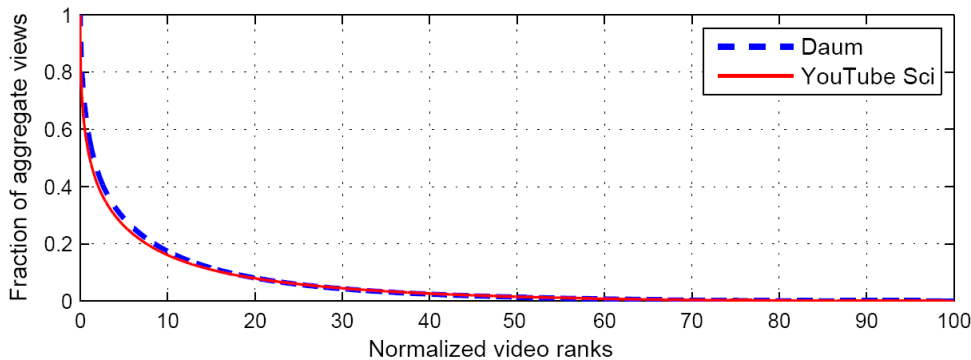


Abbildung 4: Konzentration der Aufrufe auf Videos der Kategorie "Science & Technology"

### 3.4 Lokalitäts Charakteristiken

Temporale Lokalität bedeutet, dass aus Ereignissen der jüngsten Vergangenheit auf Ereignisse der nahen Zukunft geschlossen werden kann.

#### 3.4.1 Working Set Analyse

Mit Hilfe der Working Set Analyse wird untersucht wie sich die Popularität der Videos während der Untersuchungsperiode entwickelt.

Hierzu werden die Aufgerufenen Videos der ersten Woche, der ersten 2 und 3 Wochen betrachtet. Es ist festzustellen, dass aus den in der ersten Woche aufgerufenen Videos etwa 500 der Videos auch im Rest der gesamten Beobachtungsperiode wieder aufgerufen werden. Aus den ersten 2 bzw. 3 Wochen sind es etwa 900 bzw. 1200 Videos die erneut aufgerufen werden.

Zusätzlich hierzu ist zu beobachten, dass 5-10% der an einen Tag aufgerufenen Videos auch am Folgetag aufgerufen werden. Je größer die Anfrageanzahl insgesamt, desto größer ist auch die beobachtete temporale Lokalität. Sollte der YouTube Traffic ansteigen und somit auch die Quote der temporalen Lokalität, könnten durch Caching aller Videos für die Periode eines Tages, die Auswirkungen von YouTube-Traffic auf Netzwerkressourcen minimiert werden.

Die Anzahl der innerhalb der gesamten Periode aufgerufenen Videos liegt bei 625 592, während die Anzahl der verschiedenen Videos mit 323 677, bei etwa der Hälfte liegt. Hier würde Caching etwa in der Hälfte der Fälle zu keiner weiteren Belastung der Netzwerkressourcen führen. Kritisch anzumerken bleibt jedoch, dass hierfür alle 323 677 Videos über den gesamten Zeitraum gecached werden müssten.



### 3.4.2 Globale und Lokale Popularität

Anhand der Beziehung zwischen Globaler und Lokaler Popularität bleibt festzustellen inwieweit die Aussagen die anhand der Datenaufnahme im Campusnetzwerk extrahiert wurden auch übertragbar sind auf wie globale YouTube-Anfragen.

Tatsächlich werden nur 50% der Top 100 Videos im Campusnetz aufgerufen, und an den meisten Tagen machen die global populären Videos weniger als 1% des YouTube-Traffic aus. Ursache dafür könnte sein, dass Videos durch Empfehlungen aus sozialen Netzwerken (Facebook-Links, Links aus Bekanntenkreis, etc.) bekannt gemacht werden und nicht zentral über die YouTube-Seite verteilt werden.

Offensichtlich reicht die Untersuchung der Eigenschaften der Videos aus den „Most Popular“-Listen nicht aus um daraus eindeutige Schlüsse über lokale Popularität zu ziehen. Andersherum lässt sich nicht eindeutig ausschließen, dass die lokalen Eigenschaften auf YouTube global übertrag- oder nicht übertragbar sind.

## 4 Diskussion / Schlussfolgerung

In diesem Abschnitt sollen die Erkenntnisse aus den vorhergehenden Abschnitten noch einmal kurz erwähnt und deren Auswirkungen beurteilt werden. Im Besonderen soll beurteilt werden inwiefern die Ergebnisse das verstärkte Caching von Videodaten nahe legen.

Da wie in 4. festgestellt, 98,8% des übertragenen Datenvolumens durch Videoanfragen verursacht werden, wobei etwa die Hälfte der Videos Wiederholungsanfragen sind, liegt es nahe über verschiedene an Videoanfragen angepasste Caching-Methoden nachzudenken.

Obwohl YouTube eine sehr interaktive Website ist, ist es dennoch so, dass der Großteil der Anfragen (nur Video) Get-Anfragen sind und nur ein relativ geringer Teil Post-Anfragen sind. Daraus folgt, dass das Videokontingent trotz seiner Interaktivität noch statisch genug ist, um aus Caching Vorteile zu ziehen.

Ein weiterer Punkt der hierfür spricht, ist die Videodateigröße, die deutlich über den Dateigrößen der anderen Dateitypen liegt. Diese steht in proportionalem Bezug zur Übertragungsdauer der Videodateien. Jedes gecachte Video, das nicht erneut über das Netzwerk angefordert werden muss, wirkt sich somit direkt als Entlastung der Netzwerkressourcen aus. Die Tatsache dass Videodateien im Gegensatz zu Text- oder Anwendungsdateien nicht oft modifiziert werden, und durchaus auch nach Monaten oder einem Jahr nach Posting angefragt werden, spricht für Potenziale die Video-Caching mit sich bringen kann.

Das eigentliche Problem dieser Überlegungen liegt vielmehr darin effiziente Strategien zu entwickeln, wie das Video-Caching genau umgesetzt werden soll. Denn die allgemeine Aussage, dass X % der Videoanfragen Wiederholungsanfragen sind, reicht nur um durch Cachen aller bereits angefragten Videodateien, Vorteile zu erlangen. Ob die Vorteile aus der Netzwerkressourcenentlastung, die Nachteile der Belastung der lokalen Infrastruktur aufwiegen bleibt hierbei noch zu untersuchen. Und die Frage wie lange die Videos gespeichert werden sollen ist auch nicht geklärt. Diese Fragen stellen sich nur, wenn beim Caching die zusätzlichen Meta-Daten der Videos verwendet werden sollen. Ansonsten ist die unterschiedliche Behandlung der Videos, vor der Speicherung im

Cache, gar nicht notwendig. Man speichert einfach alle angefragten Videos im Cache und entfernt immer nur die Videos die am längsten nicht angefragt wurden.

Unter 3.3 wird als Herangehensweise zu diesem Problem die Videovorkommenshäufigkeit anhand der Videopopularität abgeschätzt. Die Aufrufhäufigkeit der Videos aus dem Datenbestand zeigt einen Zipf-ähnlichen Bezug zu der Popularität der Videos. Wobei hierbei ein Rang zwischen allen abgefragten Videos im Nachhinein ermittelt wird. Dieser steht nicht direkt in Bezug zu den Ratings die zu den YouTube-Videos im Netz verfügbar sind. Erneut stellt sich die Frage wie, und anhand welcher Meta-Daten zur Laufzeit mit Hilfe einer Videoanfrage ermittelt werden soll, ob das abgefragte Video gecached werden soll oder nicht. Die Diversität der vorhandenen Videos und die daraus folgende fehlende Konzentration der Videoanfragen stellen ein weiteres Problem dar.

Die in 3.4 erwähnte temporale Lokalität sichert, dass auch die grundsätzliche Herangehensweise alle Videodateien für einen weiteren Tag(nach letzter Anfrage) zu cachen zumindest zu einer 5-10%-tigen Entlastung Netzwerkressourcen führen könnte. Ob diese Beobachtung vom Campus Netzwerk auf alle anderen Netzwerke übertragen werden kann müsste meiner Meinung nach noch weitergehend untersucht werden. Problem hierbei ist dass eine globale Untersuchung aller in YouTube vorhandenen Dateien nicht möglich ist, man aber aus den „Most Popular“-Listen extrahierte Charakteristiken nicht auf lokale Netze(z.B. Campusnetz) übertragen kann. Ausgehend von einem lokalen Netzwerk kann man jedoch auch nicht ohne weiteres auf die globalen Videocharakteristiken schließen.

Insgesamt ist zu sagen dass Video-Caching ein enormes Potential zur Netzwerkressourcenentlastung hat. Es müssen allerdings noch effiziente Strategien gefunden werden, um zwischen Videos bei denen sich Caching lohnt bzw. nicht lohnt unterscheiden zu können.

---

## Literatur

- [ 1 ] Phillipa Gill, Martin Arlitt, Zongpeng Li, Anirban Mahanti. *YouTube Traffic Characterization: A View From the Edge*, IMC 2007
- [ 2 ] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. *I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System*, IMC 2007
- [ 3 ] Business Intelligence Lowdown. Top 10 Largest Databases in the World, Feb. 2007.
- [ 4 ] M. Gittens, Kim Yong, D. Godwin, *The vital few versus the trivial many: examining the Pareto principle for software*, COMPSAC 2005. 29th Annual International
- [ 5 ] D. N. Serpanos, G. Karakostas, W.H. Wolf, *Effective Caching of Web Objects using Zipf's Law*, ICME 2000. 2000 IEEE International Conference on Multimedia and Expo2000