

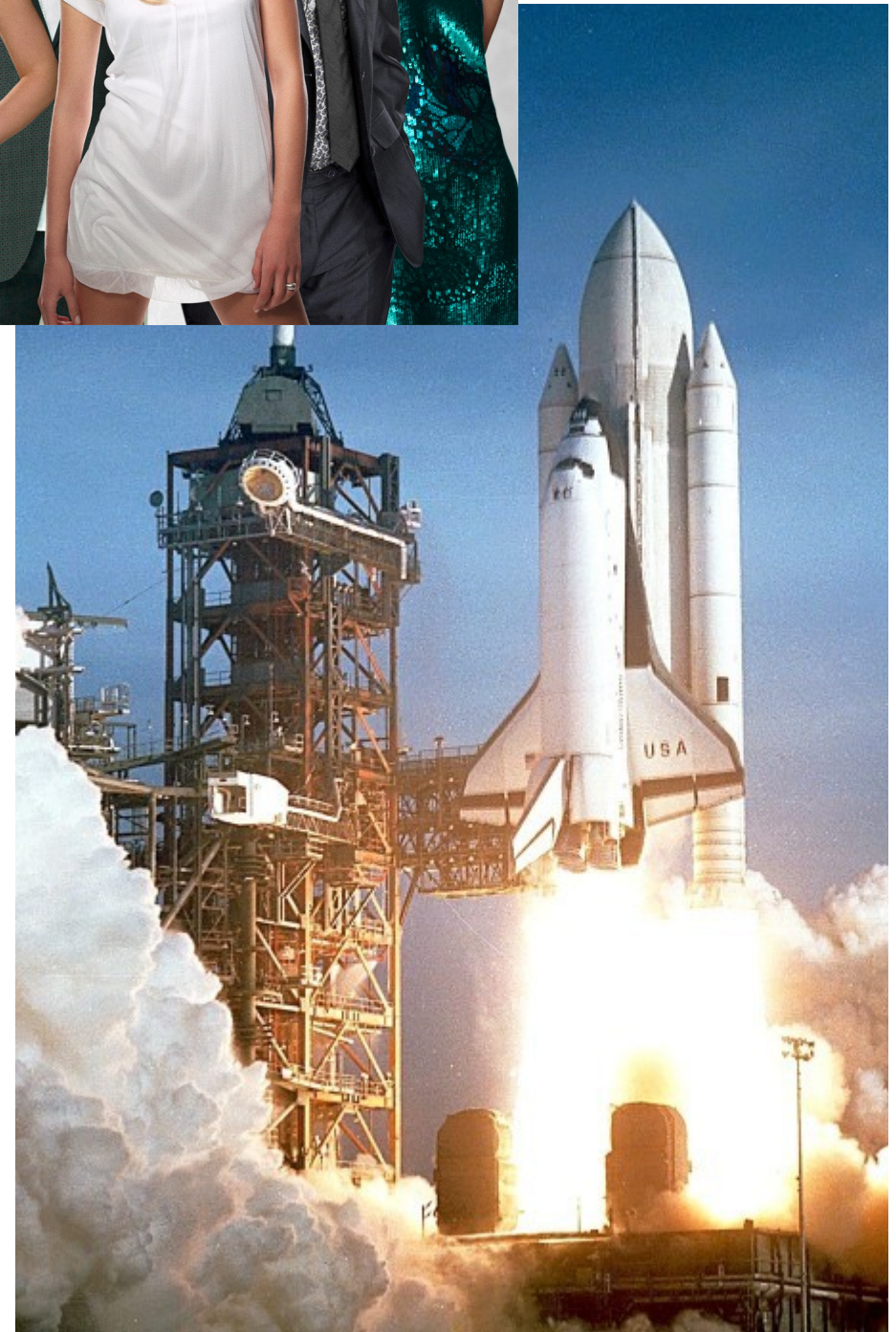


# Understanding Youtube popularity dynamics

Marian-Andrei Rizoiu

ComputationalMedia @ANU: <http://cm.cecs.anu.edu.au>

# Motivation





# Exhibit A:



V DE VEGETTA: ESPECIAL CUMPLEAÑOS: MI REGALITO (LINK DE DESCARGA)



VEGETTA777

Abonează-te 17 mil.

5.104.539 de vizionări

+ Adaugă la Trmite ... Mai multe

108.954 2.596

Statistici pentru videoclip Până pe 13 mai 2017 (inclusiv)

AFIȘĂRI	DURATA VIZIONĂRII	PE BAZĂ DE ABONAMENTE	NUMĂR DE DISTRIBUIRI
5.103.749	71 de ani	3.982	2.704

Cumulative Zilnice



# Exhibit B:



Ellie Goulding - Love Me Like You Do (Official Video)



EllieGouldingVEVO

Abonează-te 7,2 mil.

1.445.449.224 de vizionări

+ Adaugă la Trimitte ... Mai multe

5.696.268 256.469



Vaults - One Last Night (From The "Fifty Shades Of Grey" Soundtrack) [Lyric Video]



VaultsVEVO

Abonează-te 52 K

22.467.729 de vizionări

+ Adaugă la Trimitte ... Mai multe

106.146 2.356



Feb. 2015

# Exhibit B:



Ellie Goulding - Love Me Like You Do (Official Video)



EllieGouldingVEVO

Abonează-te 7,2 mil.

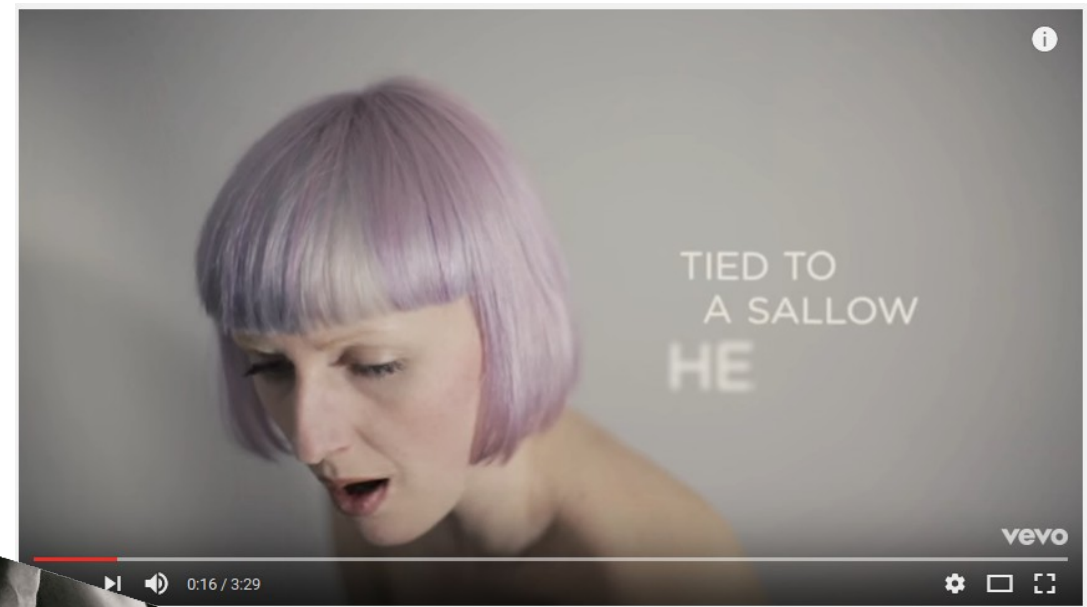
1.445.44

+ Adaugă la Trimitete Mai multe

Statistici pentru videoclip Până pe 13 mai 2017 (inclusiv)

AFIȘĂRI	DURATA VIZIONĂRII	PE BAZĂ DE ABONAMENTE	NUMĂR DE DIS
1.445.106.571	8.569 de ani	482.943	7.879.908

Cumulative Zilnice

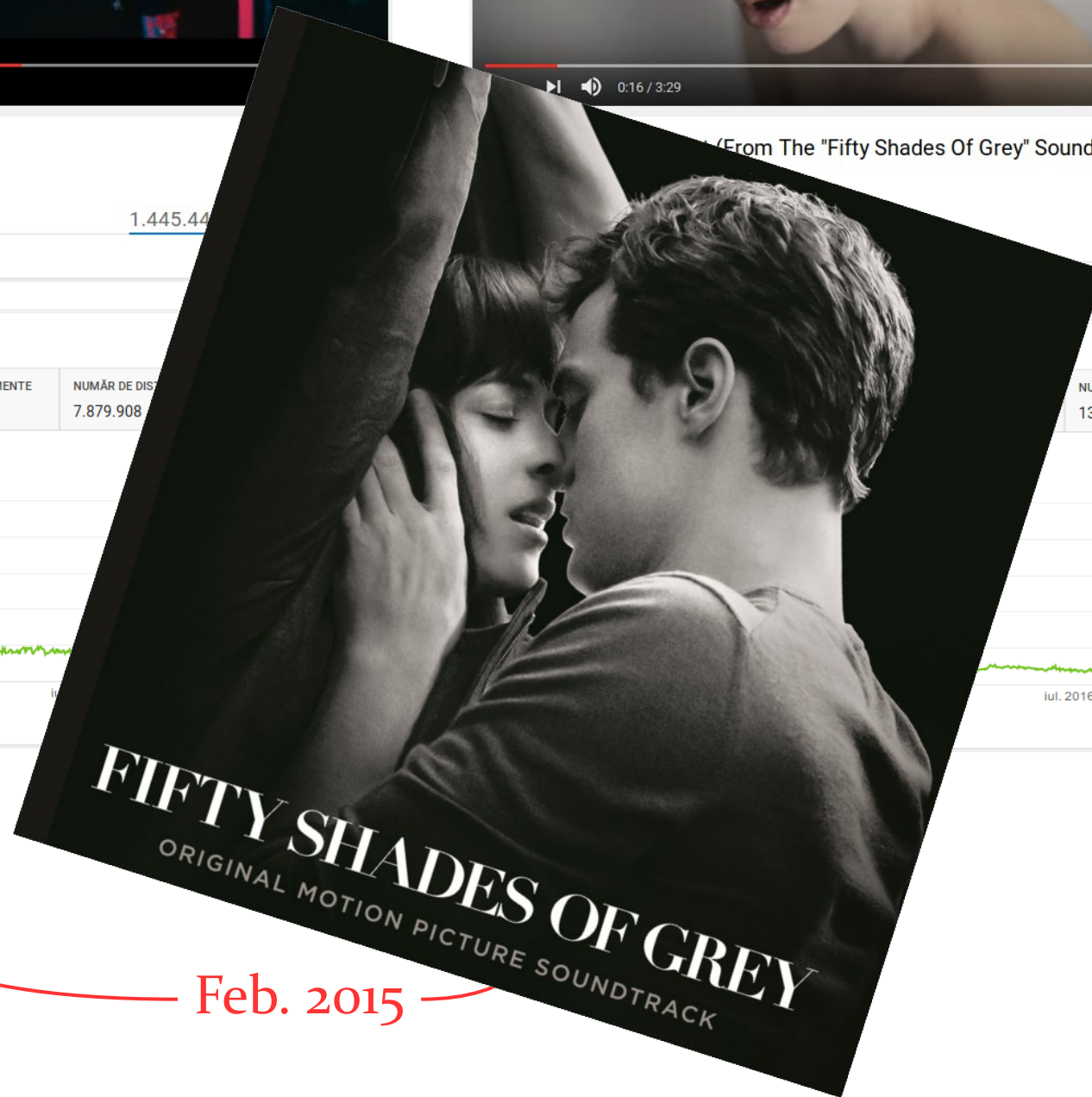


(From The "Fifty Shades Of Grey" Soundtrack)

22.467.729 de vizionări

106.146 2.356

NUMĂR DE DISTRIBUIRI
130.836



Feb. 2015



# Exhibit C:



What Narcolepsy Really Looks Like. Spoiler Alert- It Sucks.



Sleepy Sarah Elizabeth

Abonează-te 3,9 K

5.419.981 de vizionări

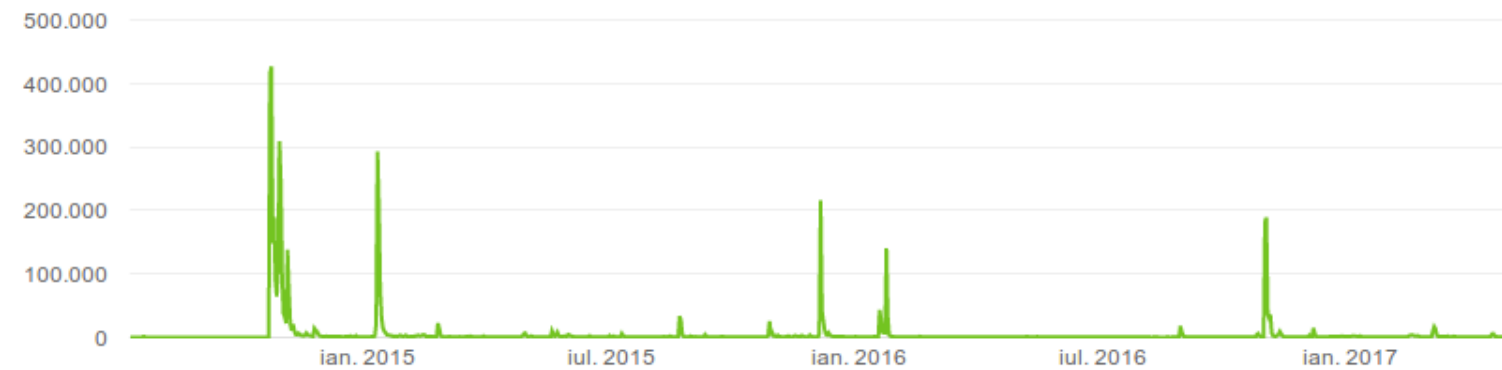
+ Adaugă la Trimite ... Mai multe

13.371 370

Statistici pentru videoclip Până pe 7 mai 2017 (inclusiv) ?

AFIȘĂRI	DURATA VIZIONĂRII	PE BAZĂ DE ABONAMENTE	NUMĂR DE DISTRIBUIRI
5.418.317	30 de ani	993	2.886

Cumulative Ziilnice ?



# Exhibit C



**Thomas Elliott**

@talexe

Brain droppings from an FY1 doctor at Broomfield Hospital. BSc in Neuroscience and Mental Health. Budding psychiatrist, if it wasn't obvious!

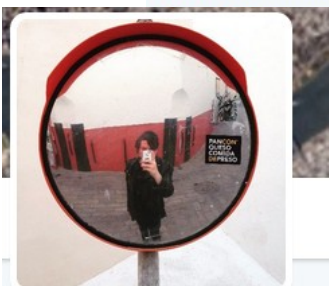
London



**MagicFlowerStone**

@MgicFlwerStne

Easter Egg Du Twitter Français - Philosophe à mes heures perdu - développeur sur @moanatar



**Adam K Olson**

@adamkolson

Designer and adoxography peddler. adam@adamkolson.com

New York, NY



**uutiset**

@8d\_mainos

Helsinki



**Women CEOs**

@CEOsWomen

Secure your .CEO domain and join a platform of over 3000 CEOs. Visit [Claim.CEO](#) Join the [Women.CEO](#) community.



What Narcolepsy Really Looks Like. Spoiler Alert- It Sucks.



Sleepy Sarah Elizabeth

Abonează-te 3,9 K

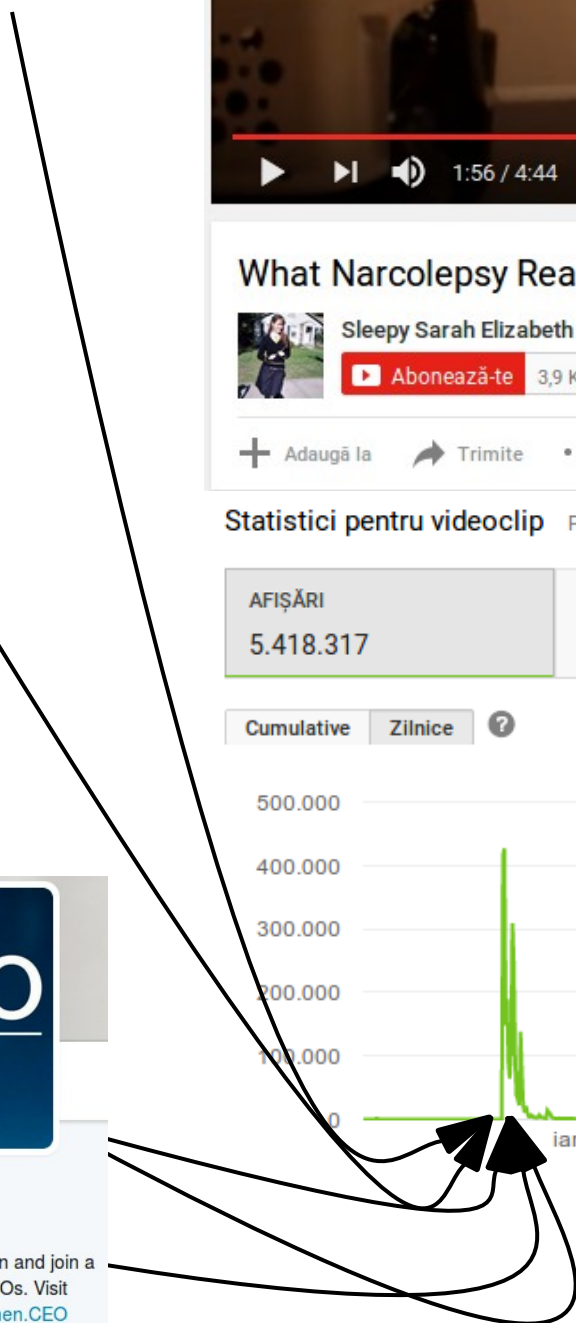
5.419.981 de vizionări

Adaugă la Trimite Mai multe

13.371 370

Statistici pentru videoclip Până pe 7 mai 2017 (inclusiv)

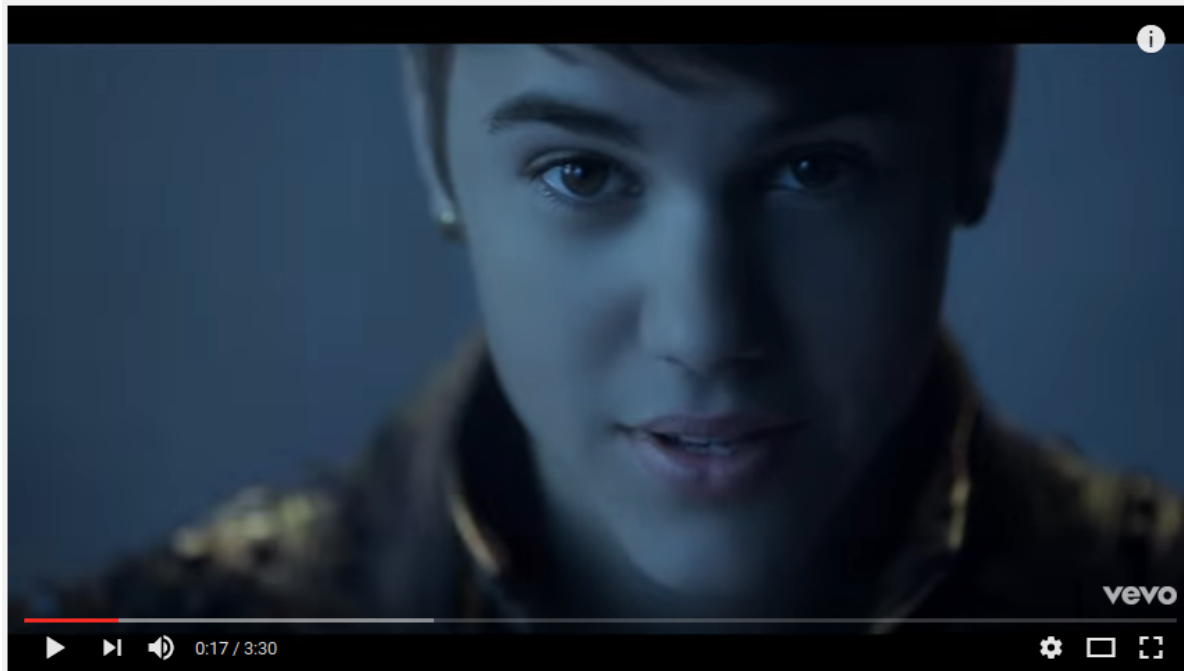
AFIȘĂRI	DURATA VIZIONĂRII	PE BAZĂ DE ABONAMENTE	NUMĂR DE DISTRIBUIRI
5.418.317	30 de ani	993	2.886



**So... external influence and promotion  
are important for popularity**



# Exhibit D:



**Justin Bieber - Boyfriend**

JustinBieberVEVO [Abonează-te](#) 29 mil.

640.560.294 de vizionări

[+](#) Adaugă la [➔](#) Trimite [⋮](#) Mai multe [👍](#) 3.183.176 [💬](#) 903.302

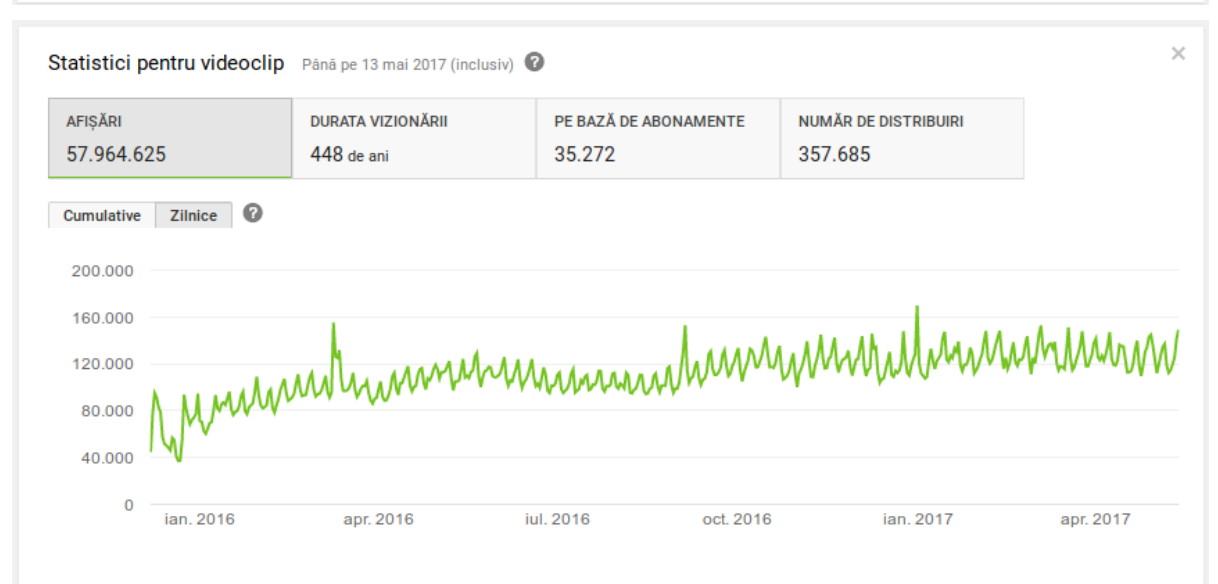


**The Beatles - Hey Jude**

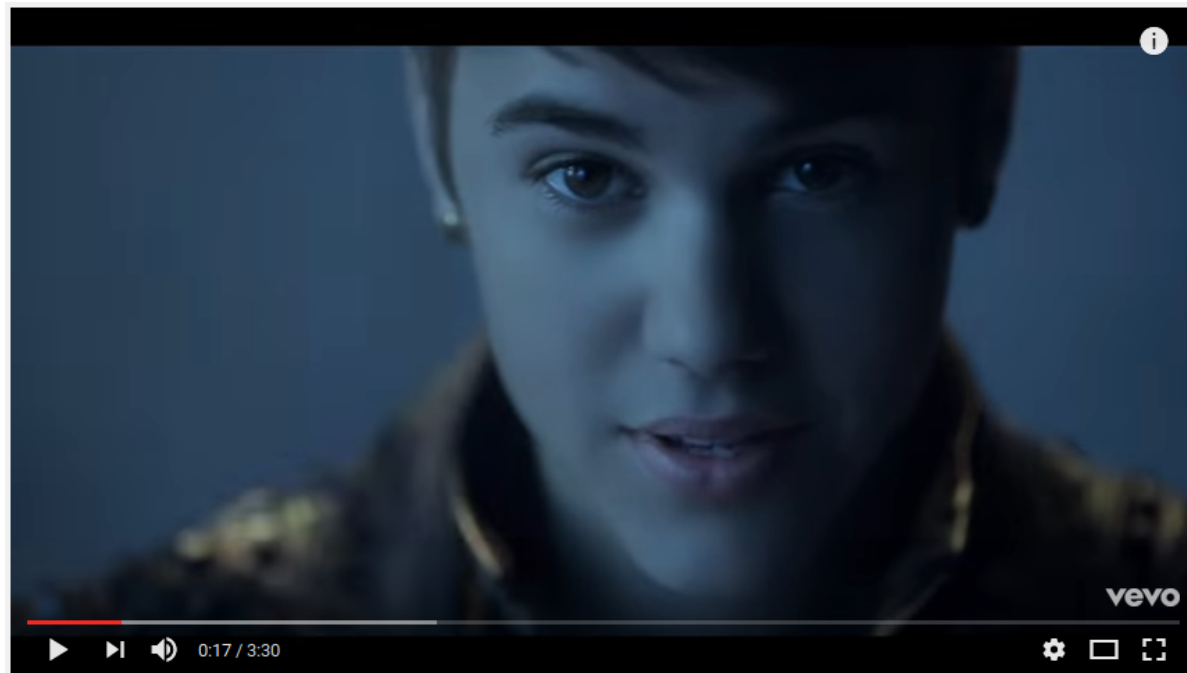
TheBeatlesVEVO [Abonează-te](#) 824 K

58.072.863 de vizionări

[+](#) Adaugă la [➔](#) Trimite [⋮](#) Mai multe [👍](#) 276.429 [💬](#) 9.368



# Exhibit D:



Justin Bieber - Boyfriend

JustinBieberVEVO Abonează-te 29 mil.

640.560.294 de vizionări

+ Adaugă la Trimitte Mai multe 3.183.176 903.302

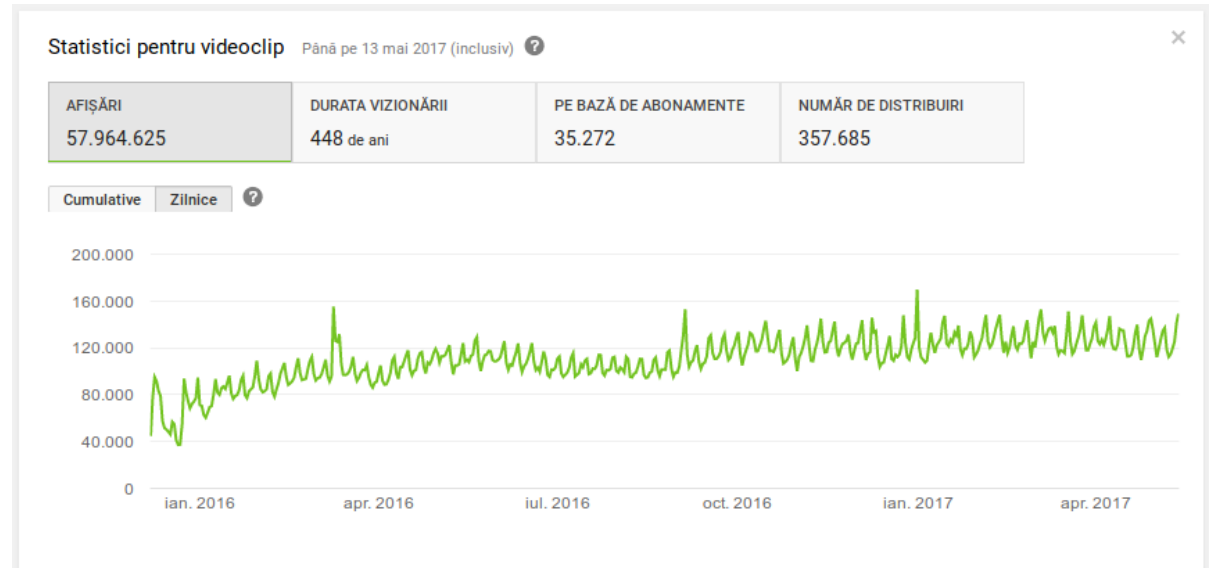


The Beatles - Hey Jude

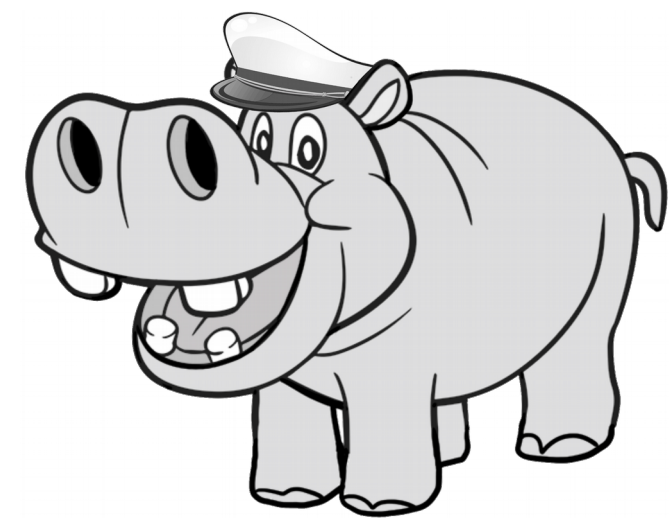
TheBeatlesVEVO Abonează-te 824 K

58.072.863 de vizionări

+ Adaugă la Trimitte Mai multe 276.429 9.368



**Videos seem to have an intrinsic quality**

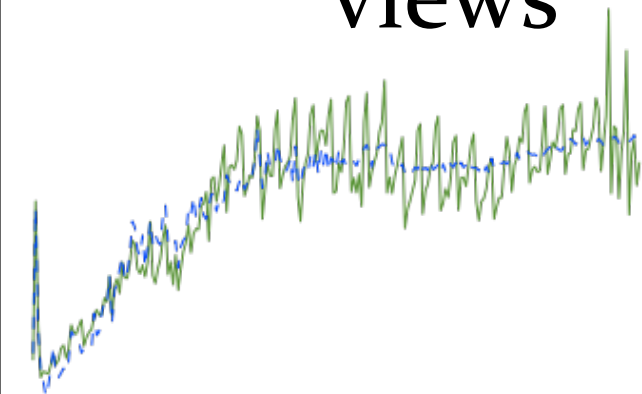
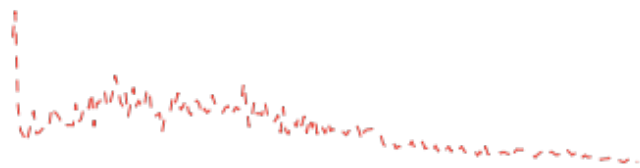


# Building HIP: Modeling and predicting popularity

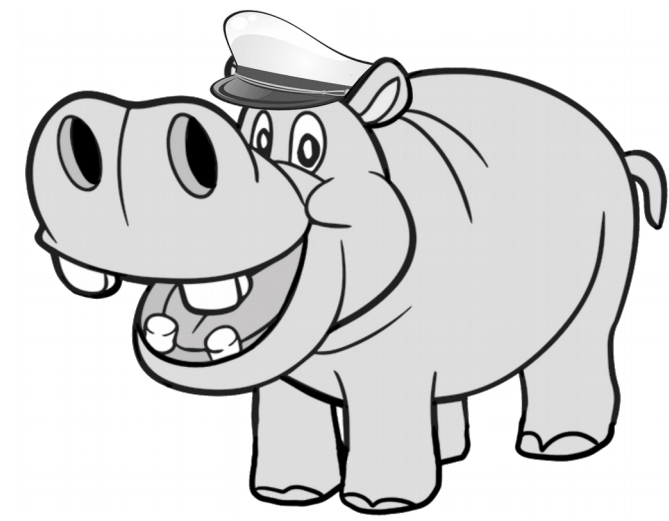
shares &  
tweets



daily  
views





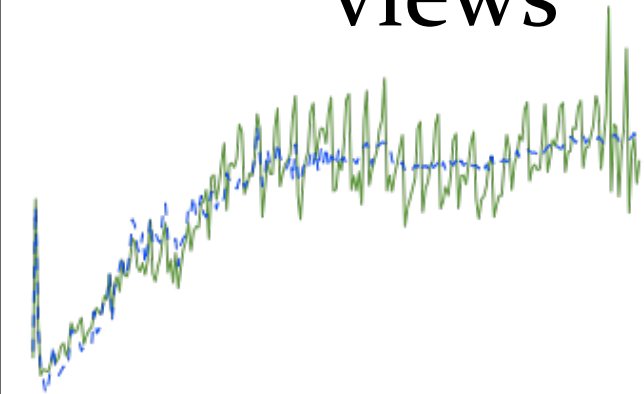
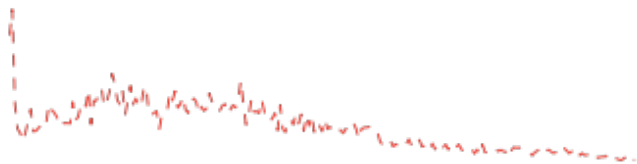


# Building HIP: Modeling and predicting popularity

shares & tweets



daily views



$\xi^u(t) := \mathbb{E}_{\mathcal{H}_t} [\lambda^u(t)]$   
 $= \mathbb{E}_{\mathcal{H}_t} \left[ \mu s(t) + \sum_{t_i < t} \phi(t - t_i) \right]$   
 $\stackrel{(a)}{=} \mu s(t) + \mathbb{E}_{t_i | N(t)} \left[ \sum_{i=1}^{N(t)} \phi(t - t_i) \right]$   
 $\stackrel{(b)}{=} \mu s(t) + \mathbb{E}_{(N(\tau), 0 < \tau < t)} \left[ \lim_{\delta \downarrow 0} \sum_{k=1}^K \mathbf{1}(dN_{k\delta}) \right]$   
 $\stackrel{(c)}{=} \mu s(t) + \lim_{\delta \downarrow 0} \mathbb{E}_{dN_{k\delta, k=1:K}} \left[ \sum_{k=1}^K \mathbf{1}(dN_{k\delta}) \right]$

$\mathbb{E}_{dN_{k\delta, k=1:K}} \left[ \sum_{k=1}^K \mathbf{1}(dN_{k\delta}) \right]$   
 $= \mathbb{E}_{dN_{1\delta}} [\mathbf{1}(dN_{1\delta} = 1) \phi(t - \delta)]$   
 $+ \mathbb{E}_{dN_{2\delta}} [\mathbf{1}(dN_{2\delta} = 1) \phi(t - 2\delta)]$   
 $+ \dots$   
 $+ \mathbb{E}_{dN_{k\delta, k=1:k'}} [\mathbf{1}(dN_{k\delta} = 1) \phi(t - k'\delta)]$   
 $+ \dots$   
 $+ \mathbb{E}_{dN_{K\delta, k=1:K}} [\mathbf{1}(dN_{K\delta} = 1) \phi(t - K\delta)]$

Each expectation term in Eq. (17) can be computed as follows.

$\mathbb{E}_{dN_{k\delta, k=1:k'}} [\mathbf{1}(dN_{k\delta} = 1) \phi(t - k'\delta)]$   
 $\stackrel{(8a)}{=} \mathbb{E}_{\tau \sim \text{Exp}(\frac{1}{\delta})} \mathbb{E}_{dN_{k'\delta} | \mathcal{H}_{(t-\tau)\delta}} [\mathbf{1}(dN_{k'\delta} = 1)] \phi(t - k'\delta)$   
 $\stackrel{(8b)}{=} \lambda^u(k'\delta) \phi(t - k'\delta)$

Furthermore, we see that  $\mathbb{E}_m[b(m)]$  can be computed in closed form, we call this modeling constant  $C$  (also defined in main text Theorem 2.1).

$\mathbb{E}_m[b(m)] = \mathbb{E}_m[\kappa m^\beta] = \kappa \int_1^\infty p(m) m^\beta dm$   
 $= \kappa \int_1^\infty (\alpha - 1) m^{-\alpha} m^\beta dm = \frac{\kappa(\alpha - 1)}{\alpha - \beta - 1} := C$  (28)

Plugging in the result of Eq. (26)–28 back to Eq. 25, we notice  $\mathbb{E}_{\mathcal{H}_{(t-\tau)\delta}}[\lambda(k'\delta)] = \xi(k'\delta)$  due to the definition of  $\xi(t)$  in Eq. (22), which yields

$\mathbb{E}_{\mathcal{H}_{k'\delta}} [\mathbf{1}(dN_{k\delta} = 1) b(m_k) \phi(t - k\delta)]_{k=k'} = C \cdot \delta \xi(k'\delta)$  (29)

Applying this result to Eq. (24) and then to Eq. 22, followed by taking the limit  $\delta \downarrow 0$ , we have:

$\xi(t) := \mathbb{E}_{\mathcal{H}_t} [\lambda(t)]$   
 $\stackrel{t=K\delta}{=} \mu s(t) + \lim_{\delta \downarrow 0} \sum_{k=1}^K C \cdot \delta \xi(k\delta) \phi(t - k\delta)$   
 $= \mu s(t) + C \int_0^t \xi(\tau) \phi(t - \tau) d\tau$   
 $\stackrel{\tau \leftarrow t - \tau}{=} \mu s(t) + C \int_0^t \xi(t - \tau) \phi(\tau) d\tau$   
 $\stackrel{\phi(\tau) = \tilde{\tau}^{-(1+\theta)}}{=} \mu s(t) + C \int_0^t \xi(t - \tau) \tilde{\tau}^{-(1+\theta)} d\tau$  (30)

Eq. (30) is Eq. (3) in the main text Theorem 2.1.

$\xi(t) := \mathbb{E}_{\mathcal{H}_t} [\lambda(t)]$   
 $= \mathbb{E}_{\mathcal{H}_t} \left[ \mu s(t) + \sum_{t_i < t} \phi_m(t - t_i) \right]$   
 $\stackrel{Eq. 12}{=} \mathbb{E}_{\mathcal{H}_t} \left[ \mu s(t) + \sum_{t_i < t} b(m_i) \phi(t - t_i) \right]$   
 $\stackrel{(12a)}{=} \mu s(t) + \mathbb{E}_{\mathcal{H}_t} \left[ \sum_{i=1}^{N(t)} b(m_i) \phi(t - t_i) \right]$

$\stackrel{(8c)}{=} \mathbb{E}_{\tau \sim \text{Exp}(\frac{1}{\delta})} \mathbb{E}_{dN_{k'\delta} | \mathcal{H}_{(t-\tau)\delta}} [\mathbf{1}(dN_{k'\delta} = 1)] \phi(t - k'\delta)$   
 $\stackrel{(8b)}{=} \lambda^u(k'\delta) \phi(t - k'\delta)$

**2.2 Computing gradients**

Eq. (34) is a non-convex objective, we use gradient-based optimization approach, and specifically L-BFGS [9] with pre-supplied gradient functions. We use the implementation supplied with the KPlot package [7] in R. We fit each video in parallel, starting with multiple random initializations to improve solution quality, and we present the solution with the lowest error function  $J$ . The gradient computations are listed as follows.

We define the error term as  $e[t] = \xi[t] - \tilde{\xi}[t]$ , Eq. (34) now becomes  $J = \frac{1}{2} \sum_{t=0}^T e^2[t]$ . Since  $\xi[t]$  are observed quantities,

$\frac{\partial e[t]}{\partial \omega} = \frac{\partial \xi[t]}{\partial \omega} - \frac{\partial \tilde{\xi}[t]}{\partial \omega}$

Here  $\text{var} \in \{\mu, \theta, C, c, \gamma, \eta\}$ . Using chain rule, we obtain:

$\frac{\partial J}{\partial \text{var}} = \sum_{t=0}^T e[t] \frac{\partial e[t]}{\partial \text{var}}$  (35)

Specifically, we compute the following partial derivatives and use them in Eq. (35) to compute the gradient.

$\frac{\partial \xi[t]}{\partial \mu} = \begin{cases} s[t] + C \sum_{\tau=1}^t \frac{\partial \xi[\tau-1]}{\partial \mu} (\tau+c)^{-(1+\theta)}, & t > 0 \\ s[0], & t = 0 \end{cases}$  (36)

for  $t > 0$ ,

$\frac{\partial \xi[t]}{\partial \theta} = C \sum_{\tau=1}^t \frac{\partial \xi[\tau-1]}{\partial \theta} (\tau+c)^{-(1+\theta)} + \xi[t-\tau] (\tau+c)^{-(1+\theta)}$   
 $= C \sum_{\tau=1}^t \left[ \frac{\partial \xi[\tau-1]}{\partial \theta} - \xi[\tau-1] \ln(\tau+c) \right] (\tau+c)^{-(1+\theta)}$  (37)

for  $t = 0$ ,  $\frac{\partial \xi[0]}{\partial \theta} = 0$ .

for  $t > 0$ ,

$\frac{\partial \xi[t]}{\partial C} = \sum_{\tau=1}^t C \frac{\partial \xi[\tau-1]}{\partial C} (\tau+c)^{-(1+\theta)} + \xi[t-\tau] (\tau+c)^{-(1+\theta)}$   
 $= \xi[t-\tau] (\tau+c)^{-(1+\theta)}$  (38)

for  $t = 0$ ,  $\frac{\partial \xi[0]}{\partial C} = 0$ .

---

$\frac{\partial \xi[t]}{\partial \omega} = \sum_{\tau=1}^t C \frac{\partial \xi[\tau-1]}{\partial \omega} (\tau+c)^{-(1+\theta)} + \xi[t-\tau] (\tau+c)^{-(1+\theta)}$  (38)

The regularizer parameters  $\omega$  is expressed as a percentage of  $J_0$  (the value of the non-normalized error function) and it is determined through a line search within

$J_{reg}(\omega, \mu, \theta, C, c) = J(\mu, \theta, C, c) + \frac{\omega}{2} \left( \left( \frac{\gamma}{\gamma_0} \right)^2 + \left( \frac{\eta}{\eta_0} \right)^2 + \left( \frac{\mu}{\mu_0} \right)^2 + \left( \frac{C}{C_0} \right)^2 \right)$  (42)

Here  $\{\gamma_0, \eta_0, \mu_0, C_0\}$  are reference values for parameters obtained by fitting the series  $\xi[t]$  without regularization. The reference values are used to normalize the parameters in the regularization process, so that they have equal weights. Intuitively using  $L^2$  normalization in square-loss is effectively putting a Gaussian prior on the parameters being regularized. We desire parameters  $c$  and  $\theta$  to take values away from zero, hence they are not regularized. The  $L^2$  regularization term is differentiable with respect with variables  $\{\gamma, \eta, \mu, C\}$  and the terms  $\frac{\partial \xi}{\partial \mu}$ ,  $\frac{\partial \xi}{\partial \theta}$ , and  $\frac{\partial \xi}{\partial C}$  are added respectively to the RHS of Eq. (40), (41), (36) and (38).

Note that the gradient computation is iterative, i.e. the computation of  $\frac{\partial \xi[t]}{\partial \omega}$  makes use of previous values in its own series  $\frac{\partial \xi[\tau]}{\partial \omega}$  for  $\tau = 1, \dots, t-1$ .

**2.3 Adding an  $L^2$  regularizer**

We add  $L^2$  regularization on the linear coefficients of the Hawkes Intensity Process to avoid overfitting. The loss function with the regularization terms are as follows.

$A_k = 1 + n + n^2 + \dots + n^k + \dots = \lim_{k \rightarrow \infty} \frac{1 - n^{k+1}}{1 - n} = \begin{cases} \frac{1}{1-n}, & n < 1 \\ \infty, & n > 1 \end{cases}$  (31)

0th capturing the endogenous property of the density model,  $A_k$  and  $n$  emphasize different in-ve chose to visualize  $A_k$  in the endo-exo map, has a direct correspondence to the sliced LTI  $w$  in main text Eq. (7) and Fig. 2(b), and a better numerical resolution for the more viral  $\omega$ , when  $n$  is close to 1. In the main text, we mates of  $A_k$  by numerically summing  $\xi[t]$  in the Eq. (7) over 10,000 discrete time steps. ad Sorrette [3] showed that the Hawkes Inten-s in a super-critical state could explain some rms of popularity observed in social media. We ver, that finite resources in the real world, such  $\propto$  human attention [10], are bound to be ex-ud online systems cannot stay indefinitely in a al regime. We argue, most online media items l by a continued interaction of exogenous stim-ogenous reaction (that may be sub- or super-adding to continued rise in popularity, or mul-s of rising and falling patterns.

Time invariance, which states that the response to a delayed input is identical and similarly delayed: if  $s(t) \rightarrow \xi(t)$  then  $s(t - t_0) \rightarrow \xi(t - t_0)$ .

We wish to show the following for Eq. (3) of the main text:

$$\xi(t - t_0) = \mu s(t - t_0) + C \int_0^{t-t_0} \tilde{\tau}^{-(1+\theta)} \xi(t - t_0 - \tau) d\tau$$

After a change of variable  $t' = t - t_0$ , we can see that the LHS is  $\xi(t')$ . For the RHS,  $\tilde{\tau}$  remains unchanged, the rest is:

$$\mu s(t') + C \int_0^{t'+t_0} \tilde{\tau}^{-(1+\theta)} \xi(t' - \tau) d\tau$$

We write the integral into two parts, i.e.,  $(0, t)$  and  $(t, t' + t_0)$ .

$$\mu s(t') + C \int_0^{t'} \tilde{\tau}^{-(1+\theta)} \xi(t' - \tau) d\tau + C \int_t^{t'+t_0} \tilde{\tau}^{-(1+\theta)} \xi(t' - \tau) d\tau$$

tion, i.e.,  $\xi(t) = 0$  for  $t'$ . The second term

$$\frac{\partial \xi[t]}{\partial c} = C \sum_{\tau=1}^t \frac{\partial \xi[\tau-1]}{\partial c} (\tau+c)^{-(1+\theta)} - (1+\theta) \xi[t-\tau] (\tau+c)^{-(1+\theta)}$$
 (39)

for  $t = 0$ ,  $\frac{\partial \xi[0]}{\partial c} = 0$ .

For the unobserved external stimuli  $\gamma$  and  $\eta$ ,

$$\frac{\partial \xi[t]}{\partial \gamma} = C \sum_{\tau=1}^t \frac{\partial \xi[\tau-1]}{\partial \gamma} (\tau+c)^{-(1+\theta)}$$
 (40)

for  $t > 0$ ,  $\frac{\partial \xi[0]}{\partial \gamma} = 1$ .

$$\frac{\partial \xi[t]}{\partial \eta} = 1 + C \sum_{\tau=1}^t \frac{\partial \xi[\tau-1]}{\partial \eta} (\tau+c)^{-(1+\theta)}$$
 (41)

for  $t = 0$ ,  $\frac{\partial \xi[0]}{\partial \eta} = 0$ .

Note that the gradient computation is iterative, i.e. the computation of  $\frac{\partial \xi[t]}{\partial \omega}$  makes use of previous values in its own series  $\frac{\partial \xi[\tau]}{\partial \omega}$  for  $\tau = 1, \dots, t-1$ .

**2.3 Adding an  $L^2$  regularizer**

We add  $L^2$  regularization on the linear coefficients of the Hawkes Intensity Process to avoid overfitting. The loss function with the regularization terms are as follows.

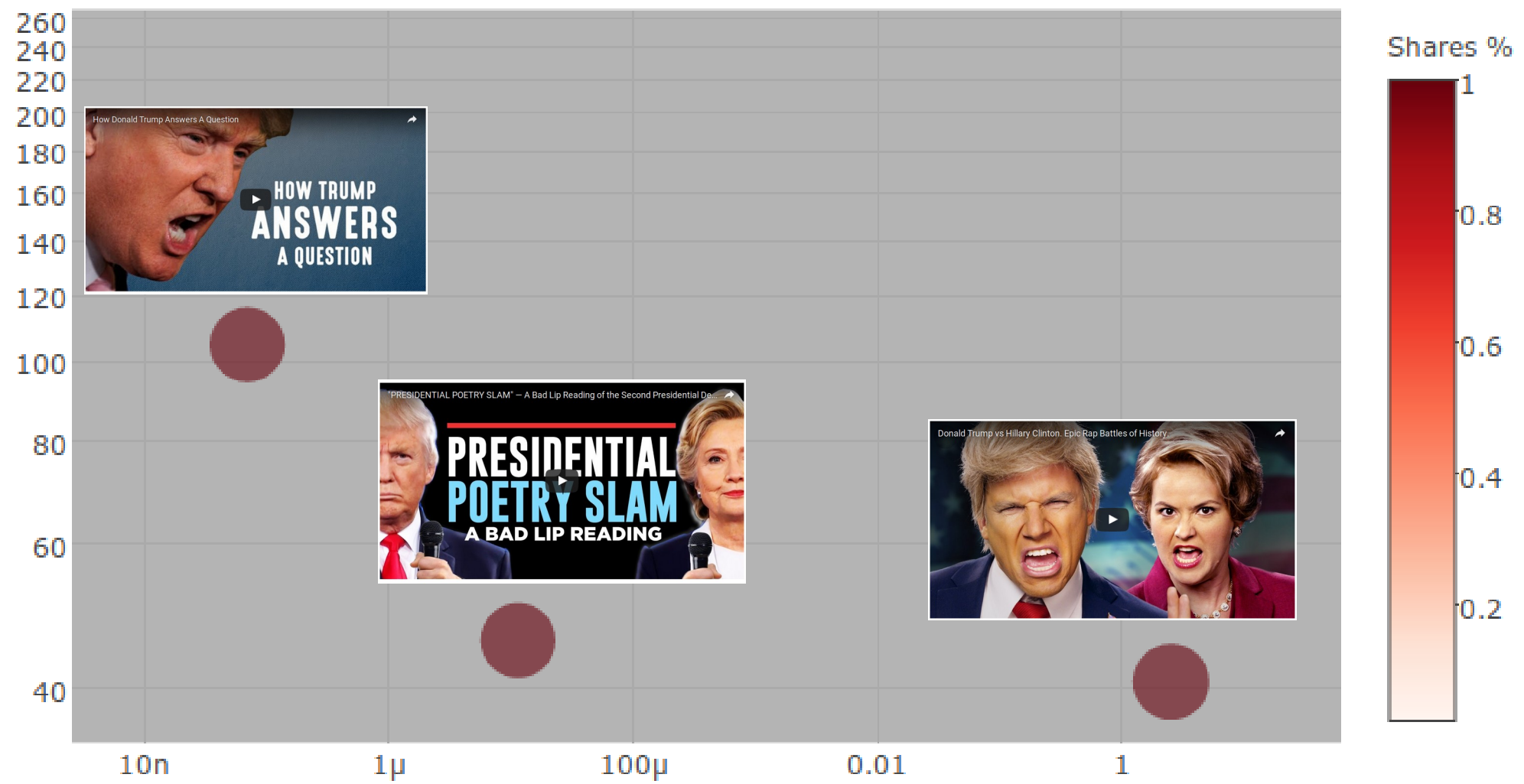
$$J_{reg}(\omega, \mu, \theta, C, c) = J(\mu, \theta, C, c) + \frac{\omega}{2} \left( \left( \frac{\gamma}{\gamma_0} \right)^2 + \left( \frac{\eta}{\eta_0} \right)^2 + \left( \frac{\mu}{\mu_0} \right)^2 + \left( \frac{C}{C_0} \right)^2 \right)$$
 (42)

Here  $\{\gamma_0, \eta_0, \mu_0, C_0\}$  are reference values for parameters obtained by fitting the series  $\xi[t]$  without regularization.

The reference values are used to normalize the parameters in the regularization process, so that they have equal weights. Intuitively using  $L^2$  normalization in square-loss is effectively putting a Gaussian prior on the parameters being regularized. We desire parameters  $c$  and  $\theta$  to take values away from zero, hence they are not regularized.

The  $L^2$  regularization term is differentiable with respect with variables  $\{\gamma, \eta, \mu, C\}$  and the terms  $\frac{\partial \xi}{\partial \mu}$ ,  $\frac{\partial \xi}{\partial \theta}$ , and  $\frac{\partial \xi}{\partial C}$  are added respectively to the RHS of Eq. (40), (41), (36) and (38).

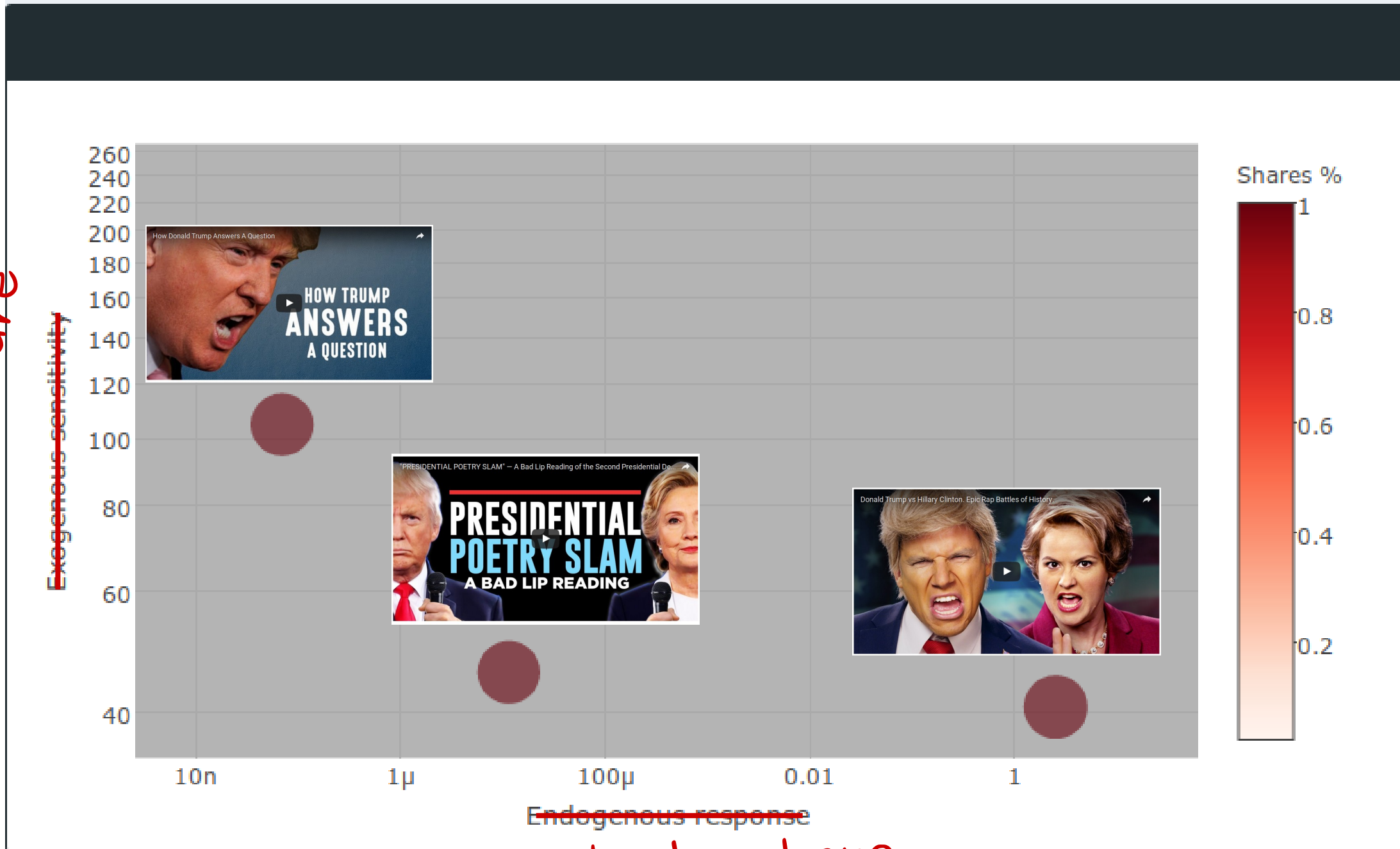
The regularizer parameters  $\omega$  is expressed as a percentage of  $J_0$  (the value of the non-normalized error function) and it is determined through a line search within



# The ~~“endo-exo”~~ map

“what do people care about”

likely to care



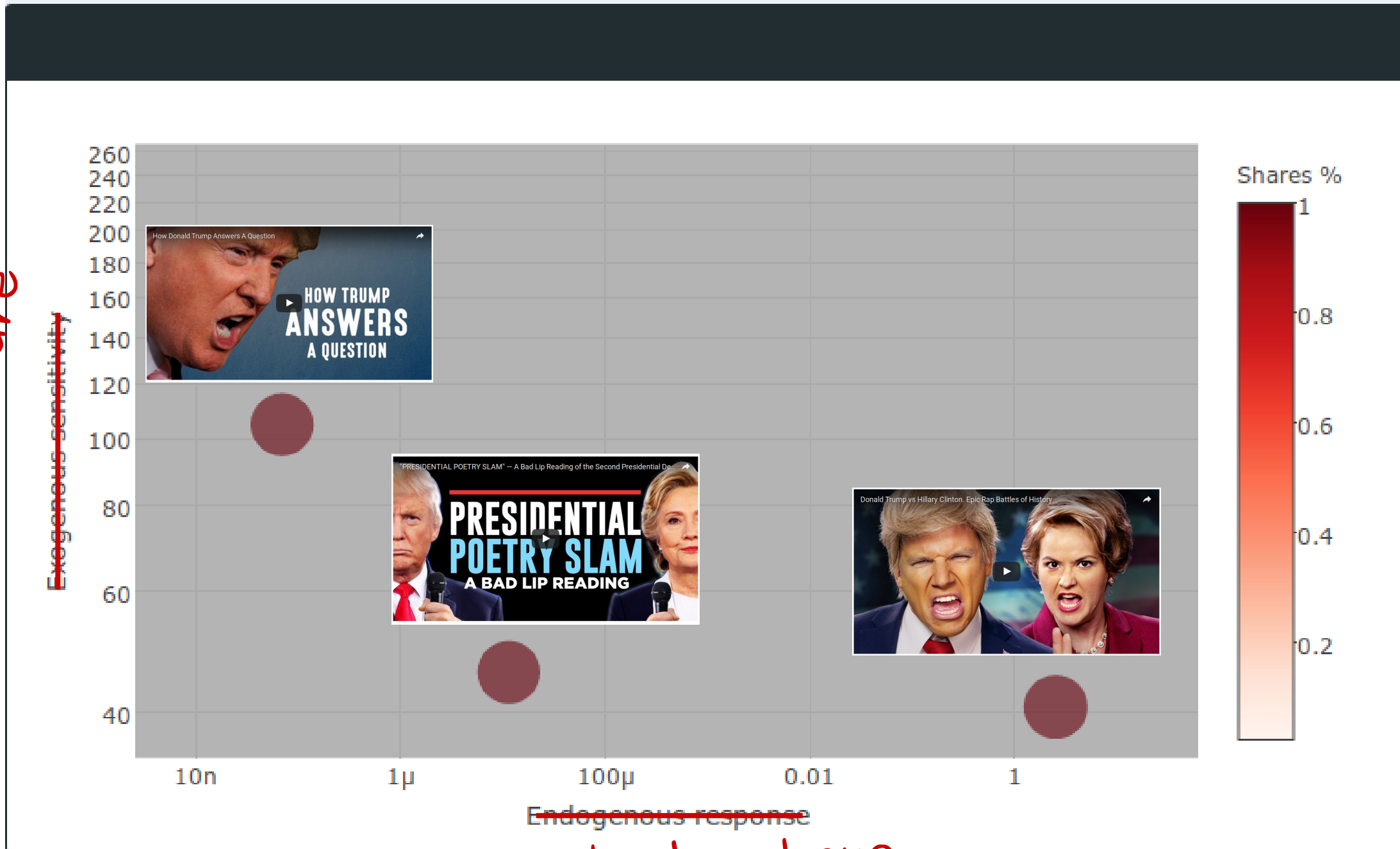
likely to share



# The ~~“endo-exo”~~ map

“what do people care about”

likely to care



likely to share

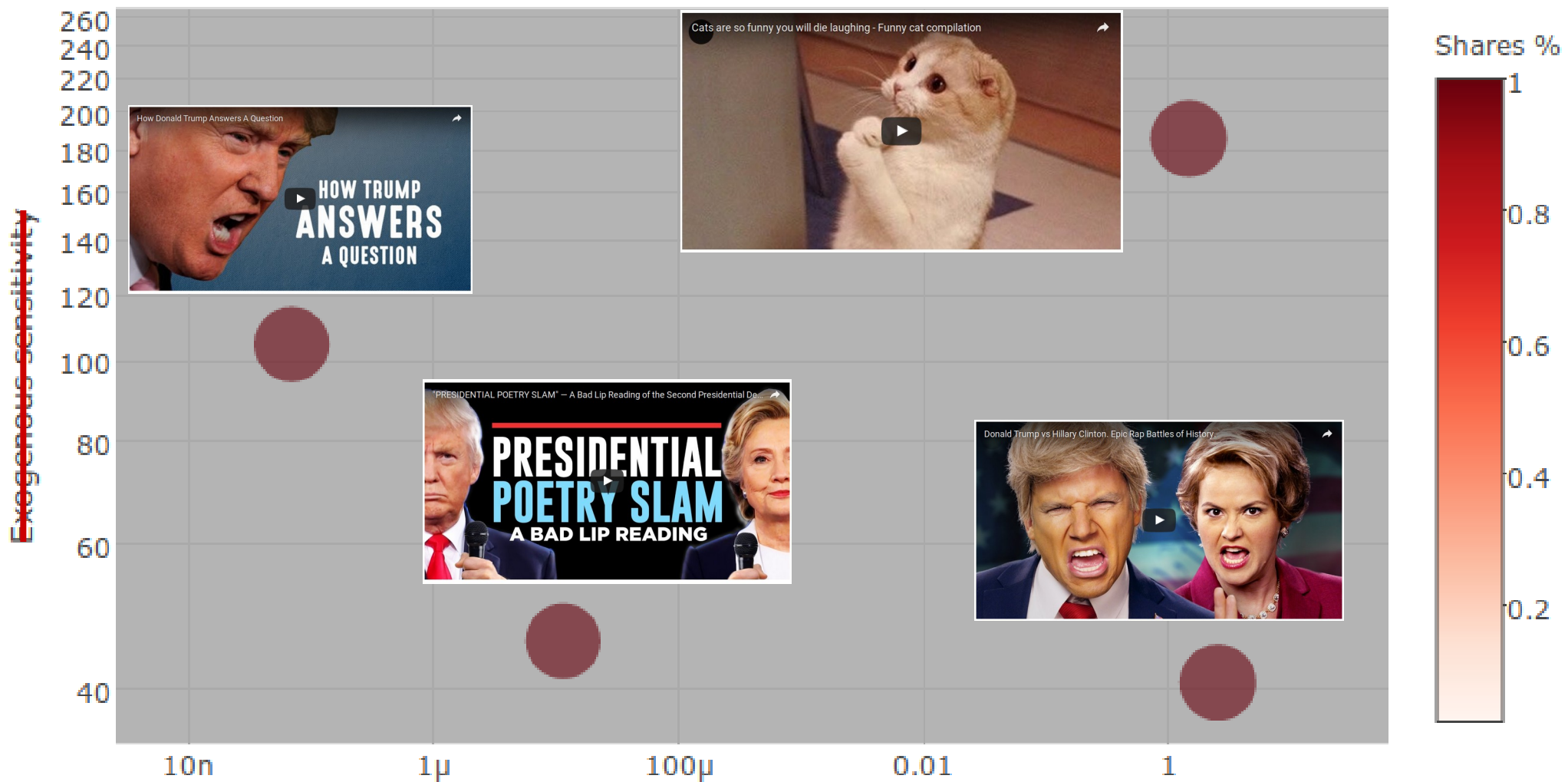
funny

serious

# The ~~“endo-exo”~~ map

“what do people care about”

likely to care



~~Endogenous response~~

likely to share

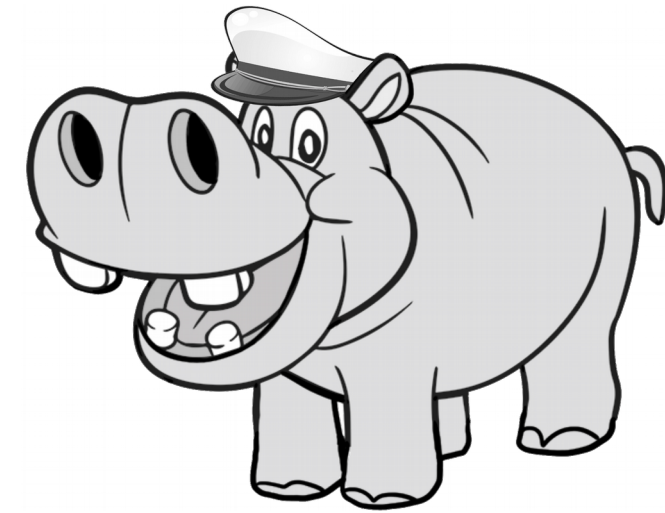
funny

serious

?!?

# Conclusion:

Principled modeling of popularity



Empirical proof that people care more about ...



than



**“cat” wars**

**news & politics**



# Thank you!

## Links:

Papers, code, dataset  
and interactive visualizer:

<https://github.com/andrei-rizoiu/hip-popularity>

Referece:

Rizolu, M.-A., Xie, L., Sanner, S., Cebrian, M., Yu, H., & Van Hentenryck, P. (2017). **Expecting to be HIP: Hawkes Intensity Processes for Social Media Popularity**. In Proceedings of the *International Conference on World Wide Web 2017*, pp. 1-9. Perth, Australia. doi: [10.1145/3038912.3052650](https://doi.org/10.1145/3038912.3052650)

[pdf at arxiv with supplementary material](#)

## HIP visualization system

This is an *interactive* visualization of the plots in the paper: the endo-exo map, observed and fitted popularity series and video metadata. It has additional visualizations of TED videos and VEVO musicians. Furthermore, it allows users to add and compare their own videos.

(access the visualizer by clicking on the thumbnail below)

