

Misogyni och främlingsfientlighet *online* – en konsekvens av anonymitet?

**EMMA
VON ESSEN
OCH JOAKIM
JANSSON**

Emma von Essen är lektor och docent vid Sociologiska institutionen vid Uppsala universitet och affilierad med Institutet för social forskning (SOFI).
emma.von.essen@uu.se

Joakim Jansson är postdoktor vid Institutet för näringslivsforskning (IFN) och affilierad med Institutet för social forskning (SOFI).
joakim.jansson@ifn.se

Sociala medier skapar möjligheter att enkelt konsumera information om politik. Samtidigt förekommer en stor mängd hat mot specifika samhällsgrupper, ofta skapade under anonymitet. Det riskerar att exkludera personer och generera felaktig information. För att förstå hur användaranonymitet påverkar hat online studerar vi en händelse där journalister oväntat tillskansade sig identiteter på en särskild grupp av användare av Flashback – ett stort online-diskussionsforum. Våra resultat visar att minskad anonymitet minskar mängden främlingsfientlighet, men inte misogyni. Även andelen länkar till fejknyheter minskade i diskussioner om invandring, men inte om kön och feminism. Resultaten antyder att främlingsfientlighet är mindre socialt accepterat än misogyni.

Den 10 September 2014 gick journalistgruppen Researchgruppen oväntat ut med att de hade identiteterna bakom användarkonton på Flashback, ett svenskt internetforum som kräver att användarna är anonyma. Snart visade det sig att det enbart rörde sig om användare som var registrerade innan mars 2007. Gruppen avslöjade även namnet på fyra användare och deras diskussionsinlägg, bl a i *Aftonbladet*. För att undersöka hur anonymitet påverkar hatiska uttalanden och falsk information i politiska diskussioner *online* betraktar vi de användare som riskerade att få sina namn och uttalanden exponerade som försöksgrupp för vilken anonymiteten minskar och de andra användarna som jämförelsegrupp.

Våra resultat visar att användare skriver färre hatiska och framför allt främlingsfientliga inlägg när de riskerar att få sitt namn och sina uttalanden exponerade, men en sådan risk tycks inte påverka antalet misogynia inlägg. Det är primärt användare som skrivit många främlingsfientliga inlägg innan händelsen som ändrar sina beteenden. Vidare ser vi att andelen främlingsfientlighet som en läsare av Flashback möter på forumet minskade, medan andelen misogyni ökade. Hat tycks även vara kopplat till felaktig information. Lägre anonymitet ger en minskning i andelen länkar till sidor med fejkade nyheter i diskussioner som rör invandring och inrikespolitik, men inte för diskussioner om kön och feminism. Kopplingen mellan anonymitet och hatiska uttalanden verkar alltså bero på vilket ämne som diskuteras. Sammantaget antyder resultaten att främlingsfientlighet är mindre socialt accepterat än misogyni.

Näthat och felaktig nyhetsinformation är ett världsomspännande samtida problem, men sambandet med användaranonymitet är inte kartlagt forskningsmässigt. Att få uttrycka sin åsikt utan att uppge namn är ett cen-

Den här artikeln är en sammanfattning av von Essen och Jansson (2020).

tralt inslag i många demokratier, t ex under röstningsprocessen i allmänna val. Dock är användaranonymitet ett tveeggat svärd som å ena sidan tillåter folk att uttrycka kontroversiella åsikter men å andra sidan omöjliggör ansvarsutkrävande gällande uttalandena (Froomkin 2017; Zhang och Kizilcec 2014). Forskning har visat att vad som skrivs och sprids *online* kan leda till politiska handlingar utanför nätet (Acemoglu m fl 2018; Qin m fl 2017; Enikolopov m fl 2018; Enikolopov m fl 2020; Zhuravskaya m fl 2020). Den här artikeln bidrar till forskningen kring hur anonymitet *online* påverkar innehållet som produceras och konsumeras (Citron 2014; Wu 2020) och hur anonymitet kan påverka individers beteende (Bursztyn m fl 2020; Hansen m fl 2018).

1. Bakgrund

Flashback är en av Sveriges mest besökta och använda internetsidor (Alexa 2017). I Sverige svarar 33 procent att de besökt Flashback, där fler av användarna är män än kvinnor (Davidsson m fl 2018). Forumet är känt, både i Sverige och internationellt, för sitt fokus på användaranonymitet och yttrandefrihet, vilket illustreras av deras slogan ”Yttrandefrihet på riktigt”. Flashback är även ökänt för att härbärgera en mängd hat, främst riktat mot invandrare, kvinnor och feminister. Forumen modereras primärt för att diskussionerna ska hålla sig till ämnet. Inga inlägg plockas bort och de är läsbara för allmänheten. Diskussionerna är organiserade i s k subforum, där ämnen alltifrån droger, rättsfall, prostituerade i Köpenhamn, politik och datorproblem, diskuteras. Vi fokuserar på diskussionerna i tre av de största subforumen som diskuterar politik – inrikespolitik, integration och invandring samt feminism. Inom varje subforum finns diskussionstrådar där specifika ämnen diskuteras som ”Nyckelord för politiska block”. Här organiseras diskussionsinläggen i kronologisk ordning.

Den 10 september 2014 gick journalistgruppen Researchgruppen ut med att de hade tillgång till identiteterna bakom tidigare anonyma konton på Flashback. Det visade sig gälla enbart användare registrerade innan mars 2007, vilket bekräftades av ägaren i en tråd på Flashback.¹ Diskussionen i tråden tyder på att händelsen skapade oro hos användarna som var registrerade innan Mars 2007. En användare som riskerade exponering skrev följande:

[...] Att för alltid vara paria när man blir googlar är illa nog, men att man riskerar att bli misshandlad eller mördad om man hängs ut är så pass allvarligt att de negativa konsekvenserna är omöjliga att greppa i dagsläget. (Flashback 2023, <https://www.flashback.org/t2526334p3>)

Journalistgruppen exponerade även namnet på fyra användare tillsammans med användarnas främlingsfientliga inlägg på forumet. Två av de fyra blev av med sina jobb och under åren som följt har fler liknande avslöjanden

¹ <https://www.flashback.org/t2526334p1>.

gjorts vilka sannolikt har använt Researchgruppens data. Händelsen var omtalad i nationell och internationell (Chen 2014) traditionell media.

2. Data

Vi inhämtade samtliga inlägg som skrivits fram t o m januari 2017 på tre av de största subforumen där politiska diskussioner förekommer – inrikespolitik, integration och invandring och feminism – och konstruerade tre mått av hat: fientlighet mot någon grupp eller person, främlingsfientlighet och misogyni. En forskningsassistent klassade totalt 4 040 slumpvis utvalda diskussionsinlägg enligt dessa mått. Exempel på ett främlingsfientligt inlägg är ”En död blatte är en bra blatte. [...]”. Ett exempel på ett inlägg som klassats som misogynt är “[...] Kvinnan är biologiskt byggd för att lyda och ta emot order [...]”. Vi använde dessa klassificeringar för att skapa tre maskininlärningsmodeller, s k *logistic lassos*. Något förenklat kan man säga att en *logistic lasso* väljer ut ord som väl predicerar hat, främlingsfientlighet respektive misogyni, baserat på de data forskningsassistenten klassat. Exempel på ord som predicerar främlingsfientlighet är *Blatt* (vilket är en sammanslagning av ord som blatte, blattar osv) medan ett ord som predicerar misogyni är *Hor* (hora, horor, hororna osv). Med dessa tre modeller skapar vi sedan mått på hat, främlingsfientlighet och misogyni för samtliga diskussionsinlägg. Utöver hat, främlingsfientlighet och misogyni har vi även information om inlägget innehöll någon länk till fejknyheter (se t ex Allcott och Gentzkow (2017) för användning av liknande mått).

För att undersöka hur anonymitet hänger ihop med näthat och länkar till fejknyheter skapade vi två paneler begränsade till tidsperioden 2012–16. Den första panelen består av användare som skrivit minst ett inlägg innan och ett inlägg efter eventet. Det betyder att det är en balanserad panel med två tidsperioder, innan och efter. Syftet med den balanserade panelen är att analysera hur användare på Flashback ändrar sitt beteende. Den andra panelen är obalanserad och består av alla inlägg. Denna panel reflekterar informationen en läsare möts av på forumet. Många privatpersoner och myndigheter använder Flashback för att inhämta information. Till exempel har Säpo och Rikskriminalpolisen använt diskussioner på forumet i jakten på terrorister och annan kriminalitet. Detta gör det relevant att undersöka det totala innehållet.

3. Resultat

Vi använder oss av en enkel modell som jämför skillnad-i-skillnader (*difference-in-differences*) för att analysera data från både vår balanserade och obalanserade panel.

$$Y_{itg} = \alpha + \beta Risk_g * Post_t + \theta Post_t + \gamma Risk_g + \varepsilon_{itg} \quad (1)$$

är utfallet för individ i (balanserade) eller inlägg i (obalanserade) under tidsperiod t tillhörande (försöksgrupp, jämförelsegrupp) g , $Risk$ indikerar om

	Försöks- grupp, före sep 2014	Försöks- grupp, efter sep 2014	Jämförelse- grupp, före sep 2014	Jämförelse- grupp, efter sep 2014
# inlägg	70,98	53,21	59,35	56,29
# hat	5,92	3,63	4,52	4,20
# främlingsfientliga	12,78	7,90	9,64	8,58
# misogyna	7,61	6,28	8,72	7,27
Std # inlägg	0,04	- 0,04	0,01	- 0,01
Std # hat	0,04	- 0,04	0,01	- 0,01
Std #främlingsfientliga	0,04	- 0,04	0,01	- 0,01
Std # misogyna	0,01	- 0,01	0,01	- 0,01
Andel hat	0,06	0,05	0,07	0,06
Andel främlingsfientliga	0,13	0,11	0,15	0,13
Andel misogyna	0,08	0,07	0,10	0,09

Tabell 1
Sammanfattande sta-
tistik per exponerad,
balanserad panel

Källa: Egen bearbetning av insamlat datamaterial.

en individ tillhör försöksgruppen, som riskerar att exponeras (registrerad innan mars 2007), eller jämförelsegruppen, och *Post* indikerar ifall observationen ägde rum innan eller efter 10 september 2014. β blir vårt estimat av förändringen i hatiska inlägg eller fejknyheter när anonymiteten minskar.

Hur ändrar användare sitt skrivande när anonymiteten minskar?

Bland användare som skapat minst ett inlägg före och efter eventet ser vi att de som riskerade exponering tydligt minskade antalet totala, hatiska och främlingsfientliga inlägg (tabell 1). Däremot minskade inte antalet misogyna inlägg. Det är inte heller någon tydlig minskning i andelen hatiska eller främlingsfientliga inlägg. För de användare som inte löpte en ökad risk att bli exponerade finns det inte lika tydliga förändringar.

Tabell 2 visar resultatet av minskad anonymitet för hatiska, främlingsfientliga och misogyna inlägg av individuella användare. Antalet hatiska inlägg minskar med 0,07 standardavvikelser och antalet främlingsfientliga inlägg minskar med 0,056 standardavvikelser bland aktiva användare som löpte en ökad risk att bli exponerade jämfört med andra. Det totala antalet inlägg minskar med liknande storlek (0,063). Däremot finns ingen förändring för antalet misogyna inlägg eller andelarna.

När vi delade upp individerna i datamaterialet utifrån om de hade en stor eller liten andel främlingsfientliga respektive misogyna inlägg visade det sig att det primärt var användare som skrev många främlingsfientliga (inte misogyna) inlägg innan händelsen som ändrar sitt beteende (von Essen och Jansson 2020).

<i>Beroende variabel:</i>							
	Std # hat	Andel hat	Std # främlingsfientliga	Andel främlingsfientliga	Std # misogyna	Andel misogyna	Std # inlägg
Post*Risk	- 0,070 (0,030)	- 0,002 (0,004)	- 0,056 (0,027)	- 0,007 (0,006)	- 0,003 (0,017)	- 0,0001 (0,005)	- 0,063 (0,024)
Post	- 0,015 (0,009)	- 0,011 (0,002)	- 0,027 (0,010)	- 0,015 (0,003)	- 0,021 (0,010)	- 0,012 (0,002)	- 0,013 (0,010)
Risk	0,035 (0,034)	- 0,006 (0,004)	0,028 (0,034)	- 0,013 (0,005)	0,002 (0,024)	- 0,021 (0,004)	0,031 (0,031)
Konstant	0,007 (0,010)	0,071 (0,001)	0,013 (0,011)	0,148 (0,002)	0,011 (0,010)	0,104 (0,002)	0,006 (0,011)
Observationer	22 318	22 318	22 318	22 318	22 318	22 318	22 318

Tabell 2
Resultat från den
balanserade panelen

Anm: Standardfelen, inom parantes, är klustrade på användarnivå (11 159 kluster).
Källa: Egen bearbetning av insamlat datamaterial.

Hur ändras den totala andelen hat och fejkheter på Flashback?

För att undersöka hur det totala innehållet en läsare av Flashback möter förändras med minskad anonymitet använder vi den obalanserade panelen. Skillnaden från den tidigare estimeringen är att vi studerar data på postnivå i stället för individnivå. Den här analysen fokuserar därmed på andelen hat, främlingsfientlighet och misogyni som Flashbackläsare möter. Vi skapade två tidsserier på veckonivå; en för användare som riskerade att bli exponerade och en för övriga användare.² Tabell 3 visar, i likhet med resultaten på individnivå, att antalet hatiska och främlingsfientliga inlägg minskar signifikant mer för de som riskerar exponering jämfört med övriga. Återigen hittar vi ingen effekt på antalet misogyna inlägg, men en minskning i det totala antalet inlägg. Detta leder till att andelen hat och främlingsfientlighet går ned, medan vi ser en ökning i andelen misogyni på forumet. Det är troligtvis en konsekvens av att det totala antalet inlägg går ner, medan antalet misogyna är konstant. För att få en bättre bild av storleken på förändringen använder vi oss av en typisk användare på Flashback som tittar på ca 84 inlägg vid ett besök. Andelen främlingsfientliga inlägg minskade med 2,4 procentenheter, vilket för en typisk besökare skulle innebära en minskning i antalet lästa främlingsfientliga inlägg från 13 till 11 inlägg (medelvärdet innan den händelse vi studerar var 16 procent). En liknande beräkning för misogyni ger en ökning från 12 till 15 misogyna inlägg som en typisk besökare ser.

² Vi använder oss här av en metod liknande den i Pettersson Lidbom och Thoursie (2013) där vi skapar en tidsserie baserat på skillnaden mellan tidsserierna för exponerade och icke-exponerade användare och sedan använder Newey-West-metoden med fyra tidsfördröjande variabler på denna tidsserie för att skatta standardfelen korrekt.

<i>Beroende variabel:</i>							
	Std # hat	Andel hat	Std # främ- lingsfientliga	Andel främ- lingsfientliga	Std # misogyna	Andel misogyna	Std # inlägg
Estimat	- 1,193 (0,185)	- 0,013 (0,003)	- 1,138 (0,176)	- 0,024 (0,004)	- 0,148 (0,223)	0,033 (0,007)	- 0,917 (0,165)
Konstant	0,556 (0,113)	0,005 (0,002)	0,530 (0,095)	0,015 (0,003)	0,069 (0,163)	- 0,043 (0,004)	0,427 (0,119)
Observationer	262	262	262	262	262	262	262

Anm: Estimeringen är baserad på en tidsserie på veckonivå över skillnaden mellan användare som riskerar exponering och de utan risk. Standardfelen, inom parantes, är skattade med Newey-West med fyra tidsfördröjande variabler.

Källa: Egen bearbetning av insamlat datamaterial.

Tabell 3
Resultat baserade på den obalanserade panelen

När vi studerar andelen inlägg med länkar till fejknyhetssidor ser vi att detta också påverkas av minskad anonymitet. Överlag ser vi en minskning i antalet länkar till fejknyhetssidor. Denna minskning är koncentrerad till subforumen inrikespolitik och integration och invandring, medan vi inte ser någon signifikant minskning inom forumet där feminism diskuteras. Detta tyder på att det i första hand är hat och främlingsfientlighet, och inte nödvändigtvis anonymitet, som är förknippat med att hänvisa till fejknyhetssidor (von Essen och Jansson 2020).

4. Avslutande diskussion

Tidigare forskning har visat att åsikter och information som sprids och konsumeras *online* kan ha en betydande inverkan på politiska handlingar (Acemoglu m fl 2018; Qin m fl 2017; Enikolopov m fl 2018, 2020; Zhuravskaya m fl 2020). Vad som sprids kan därför påverka utgången i exempelvis val. Trots detta finns det begränsat med forskning som undersöker faktorer som påverkar vad som skrivs och konsumeras. Näthat har setts som ett ökat hot mot demokratier eftersom det riskerar att tysta ned utsatta samhällsgrupper och att det kan leda till snedvridningar av den information om politiska ämnen som konsumeras. Näthatet riktar sig ofta mot specifika grupper och en betydande del av hatet skrivs utan att det går att identifiera avsändaren.

Våra resultat pekar på att minskad anonymitet leder till att aktiva användare skriver färre generellt hatiska och främlingsfientliga inlägg, men inte färre misogynna inlägg. Det totala antalet skrivna inlägg går också ner. De som hade högst andel främlingsfientlighet i inläggen innan händelsen är även de användare som minskar sin aktivitet mest. Det samma gäller *inte* de som har en hög andel misogynna inlägg. Det innebär att den typiska läsaren ser färre andel hatiska och främlingsfientliga inlägg, men en större andel misogynna inlägg. När vi tittar på andelen inlägg med länkar till fejknyheter

i samband med minskningen av anonymiteten, ser vi att dessa minskar i forum som diskuterar inrikespolitik och integration och invandring, men inte där feminism diskuteras. Det pekar på att hat, snarare än anonymitet, är kopplat till att hänvisa till fejknyheter.

Sammantaget antyder våra resultat att främlingsfientlighet är mindre socialt accepterat än misogyni. Till exempel är främlingsfientlighet, till skillnad från misogyni, enligt lag klassat som ett hatbrott. Det har också funnits en tradition bland journalister, inklusive Researchgruppen, att hänga ut politiker och övriga medborgare som uttryckt sig främlingsfientligt. Om risken för ett hårdare socialt fördömande är högre när man uttrycker sig främlingsfientligt jämfört med misogyni är det inte orimligt att användare skriver färre inlägg som är främlingsfientliga när anonymiteten minskar, medan de fortsätter skriva misogyna inlägg. För att kunna hantera hat på nätet bör framtida forskning undersöka hur möjligheten att vara anonym i olika kontexter *online* påverkar både sociala normer kopplade till främlingsfientlighet respektive misogyni och vilka typer av ämnen och politiska åskådningar som diskuteras.

REFERENSER

- Acemoglu, D, T A Hassan och A Tahoun (2018), "The Power of the Street: Evidence from Egypt's Arab Spring", *The Review of Financial Studies*, vol 31, s 1–42.
- Alexa (2017), "Internet Rankings", <https://www.alexa.com/siteinfo/flashback.org> [nedladdad 2017-01-01].
- Allcott, H och M Gentzkow (2017), "Social Media and Fake News in the 2016 Election", *Journal of Economic Perspectives*, vol 31, s 211–236.
- Bursztyn, L, G Egorov och S Fiorin (2020), "From Extreme to Mainstream: The Erosion of Social Norms", *American Economic Review*, vol 110, s 3522–3548.
- Chen (2014), "The Troll Hunters", *MIT Technological Review*, 18 december 2014, <https://www.technologyreview.com/s/533426/the-troll-hunters/>.
- Citron, D K (2014), *Hate Crimes in Cyberspace*, Harvard University Press, Cambridge MA.
- Cohen, K, F Johansson, L Kaati och J C Mork (2014), "Detecting Linguistic Markers for Radical Violence in Social Media", *Terrorism and Political Violence*, vol 26, s 246–256.
- Davidsson, P, M Palm och A M Mandre (2018), *Svenskarna och Internet 2018*, IIS (Internetstiftelsen i Sverige), Stockholm.
- Enikolopov, R, A Makarin och M Petrova (2020), "Social Media and Protest Participation: Evidence from Russia", *Econometrica*, vol 4, s 1479–1514.
- Enikolopov, R, M Petrova och K Sonin (2018), "Social Media and Corruption", *American Economic Journal: Applied Economics*, vol 10, s 150–174.
- von Essen, E och J Jansson (2020), "Misogynistic and Xenophobic Hate Language Online: A Matter of Anonymity", IFN Working Paper 1350, Stockholm. Se även senaste utkastet som finns tillgängligt på Joakim Janssons hemsida: <https://sites.google.com/site/joakimjanssoneconomist>.
- Flashback (2023), "Researchgruppen påstår att de har Flashbacks databas (del 2)", <https://www.flashback.org/t2526334p3>.
- Froomkin, A M (2017), "Lessons Learned too Well: Anonymity in a Time of Surveillance", *Arizona Law Review*, vol 59, s 95–160.
- Hansen, S, M McMahon och A Prat (2018), "Transparency and Deliberation within the FOMC: A Computational Linguistics Approach", *Quarterly Journal of Economics*, vol 133, s 801–870.
- Pettersson Lidbom, P och P S Thoursie (2013), "Temporary Disability Insurance and Labor Supply: Evidence from a Natural Experiment", *The Scandinavian Journal of Economics*, vol 115, s 485–507.
- Qin, B, D Strömberg, och Y Wu (2017), "Why Does China Allow Freer Social Media? Protests versus Surveillance and Propaganda", *Journal of Economic Perspectives*, vol 31, s 117–140.

Wu, A H (2020), "Gender Bias among Professionals: An Identity-based Interpretation", *Review of Economics and Statistics*, vol 102, s 867–880.

Zhang, K och R F Kizilcec (2014), "Anonymity in Social Media: Effects of Content Controversiality and Social Endorsement on Sharing Behavior", i *Eighth International AAAI Conference on Weblogs and Social Media*, 1–4 juni 2014, Ann Arbor, Michigan.

Zhuravskaya, E, M Petrova och R Enikolopov (2020), "Political Effects of the Internet and Social Media", *Annual Review of Economics*, vol 12, s 415–438.