

Bisbilhotando vídeos ao vivo alheios do YouTube: observações, *insights* e caminhos aplicados a Seminários de Informática do ForPPGC-PR em 2023-2

Josiney de Souza^{1,2} Marcos S. Sunye²

¹IFC *campus* Brusque
Av. Hugo Schlosser, 605 – Jardim Maluche, Brusque – SC, 88354-300

²Departamento de Informática da UFPR
Av. Cel. Francisco H. dos Santos, 100 – Jardim das Américas, Curitiba – PR, 81530-000

josiney.souza@ifc.edu.br, sunye@inf.ufpr.br

Abstract. *Currently, in the Internet Revolution, physical relationships are taken to cyberspace, such as social media. Its objectives are studied and analyzed by metrics, such as popularity. On YouTube, despite existing metrics, they are only accessible to channel administrators; other users must rely on other ways. This work proposes a simple system for monitoring live YouTube videos by any user. The system is applied to second academic semester in 2023 of Forum of Postgraduate Programs in Computing in Paraná. It is observed that viewers begin to leave the broadcast after about an hour and there is a decrease in the live audience throughout the semester. Finally, possibilities for future work are discussed.*

Resumo. *Atualmente, na Revolução da Internet, as relações físicas são levadas ao ciberespaço, como as mídias sociais. Seus objetivos, são estudados e analisados por métricas, como a popularidade. No YouTube, apesar das métricas existentes, só são acessíveis aos administradores do canal; demais usuários devem recorrer a outros meios. Este trabalho propõe um sistema simples de monitoramento de vídeos ao vivo do YouTube por qualquer usuário. O sistema é aplicado ao semestre letivo 2023-2 do Fórum dos Programas de Pós-Graduação em Computação do Paraná. Observa-se que os espectadores começam a deixar a transmissão após cerca de uma hora e há uma diminuição do público ao vivo ao longo do semestre. Por fim, se discute possibilidades de trabalhos futuros.*

1. Introdução

Atualmente, se vive a Revolução da Internet, onde as diversas relações físicas da sociedade (trabalho, educação, entretenimento, relacionamentos, etc) em muitos casos se misturam com as relações virtuais permitidas pela rede mundial de computadores. Esse ciberespaço criado é então ocupado pelo ser humano digital, que se organiza em sistemas colaborativos para desempenhar trabalhos em grupo [NICOLACI-DA-COSTA and PIMENTEL 2011].

Com objetivos diversos como armazenar e trocar experiências ou gerenciar o conhecimento, até mesmo as redes sociais têm aderido ao ciberespaço, sendo chamadas de mídias sociais [de Lemos MEIRA et al. 2011]. E, para saber se seus objetivos estão sendo

alcançados, faz-se necessário a análise de seu uso a partir de métricas, como centralidade, popularidade e densidade.

Dentre as mídias sociais, os blogs de quaisquer categorias (texto, fotos, vídeos) são ao mesmo tempo um meio de comunicação, por potencialmente alcançar muitas pessoas, e um sistema colaborativo, por permitir a conversação colaboração de usuários sobre uma postagem [PIMENTEL et al. 2011]. Nesse contexto, tomando como base o YouTube, uma das mídias sociais mais usadas no Brasil [Souza 2023], a popularidade é uma das principais métricas da mídia social.

Apesar de possuir uma ferramenta de análise do vídeo, que permite ver o número total de visualizações de vídeos, mensagens no *chat* e espectadores simultâneos ao longo do tempo, algumas dessas métricas são de acesso apenas dos proprietários do canal, como o número de espectadores em determinados momentos do vídeo e novas visualizações.

Devido a essa característica, perguntas “quantos espectadores assistem a uma transmissão ao vivo?”, “qual o comportamento do espectador ao longo da transmissão?”, “qual a relação entre o número de visualizações ao vivo e o total de visualizações dos vídeos?” e “o número de visualizações ao vivo reflete a lista de presença assinada?” podem ocorrer em qualquer espectador de vídeos no YouTube durante uma transmissão.

Para tentar responder a essas perguntas, este trabalho fez um sistema de monitoramento de vídeos ao vivo do YouTube que obtém o número de espectadores simultâneos a cada minuto de transmissão e os guarda em um banco de dados para posterior tratamento e contagem estatística sobre essa métrica.

O teste do sistema foi realizado sobre as transmissões ao vivo do Fórum de Programas de Pós-Graduação em Computação do Paraná (ForPPGC-PR) realizadas no segundo semestre letivo de 2023. Apesar das transmissões e do semestre ainda não terem finalizado, é possível verificar comportamentos e padrões dos espectadores do ForPPGC-PR.

As demais seção estão assim organizadas: na Seção 2, apresenta-se o ForPPGC-PR; na Seção 3, mostra-se o sistema de monitoramento; na Seção 4, estão as observações, os resultados obtidos e discussões sobre as transmissões do semestre atual do ForPPGC-PR após monitoramento realizado; na Seção 5, faz-se um fechamento deste trabalho; na Seção 6 estão agradecimentos, seguido pelas referências utilizadas.

2. Fórum de Programas de Pós-Graduação em Computação do Paraná

O Fórum de Programas de Pós-Graduação em Computação do Paraná (ForPPGC-PR), criado em 2019 e que ocorre anualmente, é uma iniciativa importante para a formação dos estudantes dos diversos programas de Pós-Graduação do Estado; em um trabalho em conjunto das diferentes Instituições de Ensino Superior (IES) que compõem a iniciativa.

Além dos eventos anuais, uma das ações que fomenta discussões na área de Computação e auxilia a formação de estudantes são as transmissões realizadas em canal do YouTube. Muitas vezes, essas transmissões são aproveitadas como conteúdo dos componentes curriculares dos Programas de Pós-Graduação, a exemplo de Seminários em Informática I e Seminários em Informática II do Programa de Pós-Graduação em Informática (PPGInf) da Universidade Federal do Paraná (UFPR) que contam como créditos acadêmicos para mestrado e doutorado, respectivamente.

Considerando o segundo semestre de 2023, toda terça-feira, das 19h às 21h, há uma transmissão diferente que conta como aula para os estudantes matriculados nesses componentes curriculares. Para compor a presença, ao fim da apresentação há um formulário on-line que o estudante preenche com seus dados pessoais e informações referente à apresentação assistida.

3. O sistema de monitoramento

O sistema de monitoramento é composto por duas partes. A primeira, é um *script* que executa três passos: (i) obtém a página do YouTube com o vídeo ao vivo do dia usando a ferramenta `wget(1)`; (ii) recorta a informação de “assistindo agora” da página baixada com `grep(1)`; e (iii) escreve o dado em arquivo CSV no formato “HORA;VISUALIZADORES”. O código abaixo é o *script* “captura.sh” criado:

```
#!/bin/bash

DATA_DO_DIA=$(date +%Y%m%d)
HORA_DO_DIA=$(date +%H%M)
HORA_DO_DIA_FORM=$(date +%H:%M)
ENDERECO="https://www.youtube.com/watch?v=54NmqcpiLlk"

cd ~/desenvolvimento/git/seminarios

mkdir -p ${DATA_DO_DIA}

wget ${ENDERECO} -O ${DATA_DO_DIA}/${HORA_DO_DIA}.html

# VISU=$(grep "views" $1 | \
VISU=$(grep "views" ${DATA_DO_DIA}/${HORA_DO_DIA}.html | \
sed -e "s/.*\ (views.*assistindo agora)\ .*\/\1/" | \
cut -d' ' -f7)

echo "${HORA_DO_DIA_FORM};${VISU}" >>
    ${DATA_DO_DIA}/${DATA_DO_DIA}.csv

exit 0
```

A segunda parte é colocar o *script* para execução automática no horário dos seminários. A ferramenta utilizada para esse agendamento de tarefas é o `crontab(1)`. No código ilustrativo abaixo, agenda-se a execução do *script* “captura.sh” para todo minuto entre 19h00 e 21h59 de toda e qualquer terça-feira do mês:

```
# m h dom mon dow command
* 19-21 * * 2 captura.sh
```

4. Resultados e discussões

As transmissões dos seminários ao vivo do ForPPGC-PR no semestre atual começaram em primeiro de agosto porém o sistema de captura só foi criado no fim desse mês e

só esteve disponível para uso nas transmissões a partir de setembro. Assim, a primeira transmissão monitorada foi a de 05 de setembro de 2023, do vídeo intitulado “Robôs autônomos, Inteligência artificial e competições de robótica, João Alberto Faibro (UTFPR)” (<https://www.youtube.com/watch?v=s-4KHWSrbK4>), resumida conforme Figura 1 abaixo:

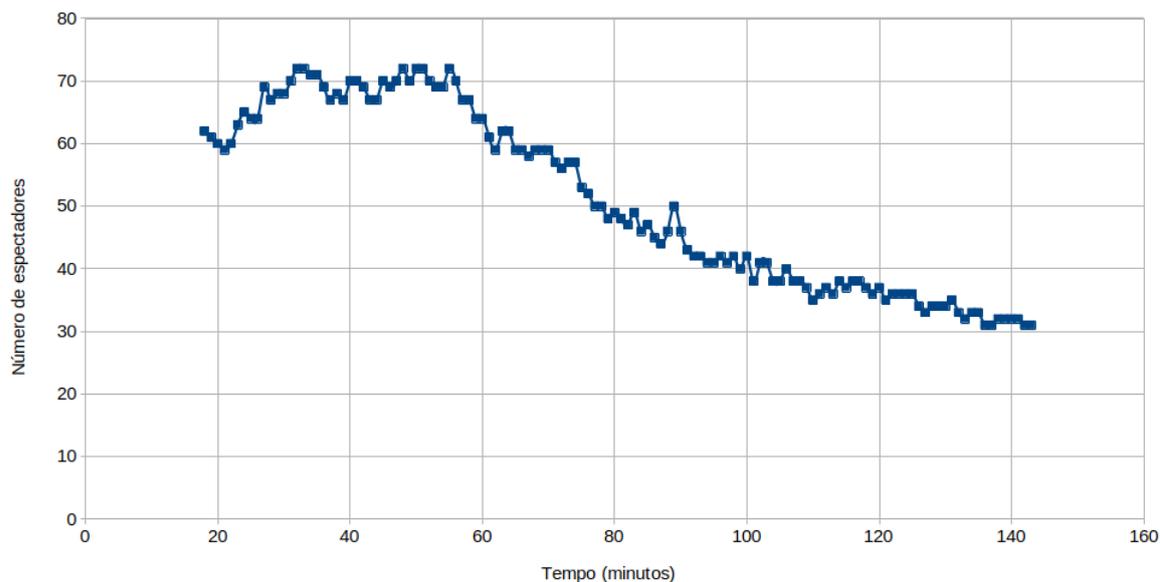


Figura 1. Número de espectadores simultâneos ao vivo ao longo do tempo para a primeira transmissão monitorada do ForPPGC-PR, em 05/09/2023

A partir dos dados coletados, notou-se que a primeira transmissão teve um máximo e um mínimo de 72 e 31 espectadores simultâneos, respectivamente; além de média 51,08, desvio padrão 14,22; mediana 49 e moda 38 aparecendo como oito resultados, ou seja, representando oito dos 146 minutos transmitidos. Embora se tenha perdido os dados dos primeiros minutos da transmissão, ainda assim é possível verificar que, após aproximadamente uma hora (60 minutos), o número de espectadores começa a reduzir.

Com essas observações, foi possível ter uma primeira hipótese e parâmetros de base para a resposta das duas primeiras perguntas levantadas (“quantos espectadores assistem a uma transmissão ao vivo?” e “qual o comportamento do espectador ao longo da transmissão?”) mas ainda era necessário observar as demais transmissões para comparação. Assim, foram monitoradas as demais transmissões de setembro conforme Figura 2 abaixo:

Nota-se que em duas das quatro transmissões de setembro os espectadores começaram a deixar a transmissão a partir de uma hora (dias 05/09 em azul e 26/09 em verde). Ou seja, a hipótese poderia ser atualizada para dizer que metade das transmissões consideram saídas após uma hora. A outra metade das transmissões tiveram saídas após uma hora e meia (90 minutos) aproximadamente. O mês de outubro também se comportou de forma semelhante, conforme Figura 3 abaixo:

Observa-se que a transmissão de 10/10, em vermelho, embora tenha se iniciado às 19h30 (30 minutos após às 19h), começou a demonstrar um declínio mais acentuado após aproximadamente 90 minutos; totalizando o declínio maior após uma hora do início

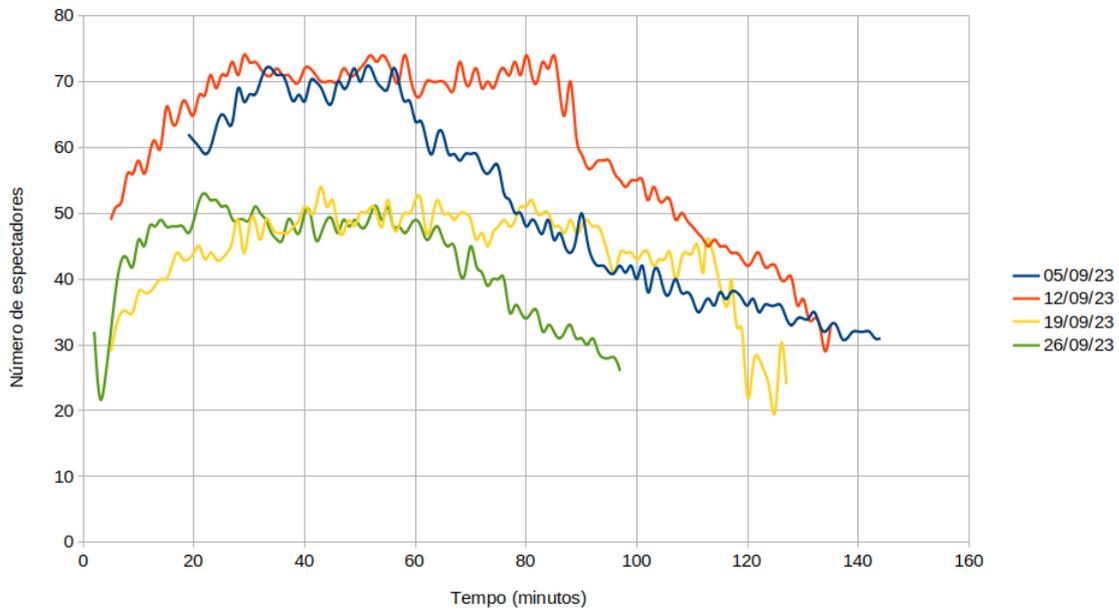


Figura 2. Número de espectadores simultâneos ao vivo ao longo do tempo das transmissões monitoradas do ForPPGC-PR em setembro de 2023

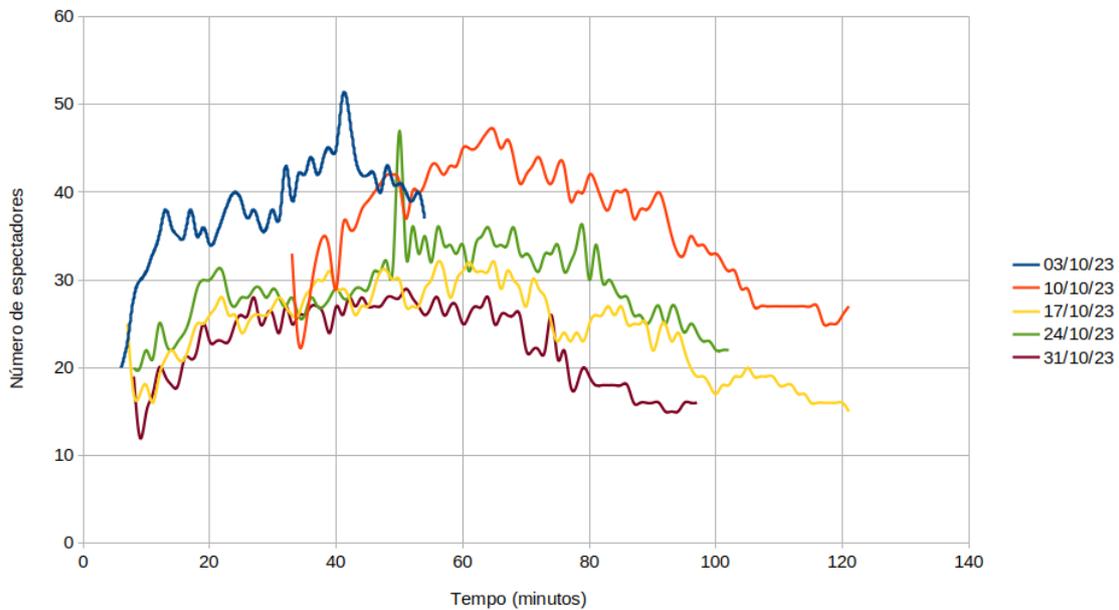


Figura 3. Número de espectadores simultâneos ao vivo ao longo do tempo das transmissões monitoradas do ForPPGC-PR em outubro de 2023

da transmissão. As transmissões de 17/10, em amarelo, e de 31/10, em vermelho escuro, salvo alguns picos isolados de aumentos de espectadores, demonstraram uma tendência de queda após aproximadamente 70 minutos do início da transmissão mas, devido a iniciarem próximas das 19h10, preservam a hipótese de diminuição após aproximadamente uma hora do início da transmissão em aproximadamente metade delas.

Outras observações dizem respeito à média de público ao vivo ao longo do se-

mestre. De todos os Programas de Pós-Graduação do Estado, há 76 matriculados; destes, 26 do Programa de Pós-Graduação em Informática da UFPR (PPGInf). Para confrontar esses, conforme representado na Tabela 1 e corroborado pela Figura 4, a presença dos espectadores têm diminuído com o passar do semestre.

Tabela 1. Média de espectadores ao vivo nas transmissões do ForPPGC-PR em setembro e outubro de 2023

Setembro	Data	05/09/23	12/09/23	19/09/23	26/09/23	
	Média	51,0873	60,9924	44,3554	42,9375	
Outubro	Data	03/10/23	10/10/23	17/10/23	24/10/23	31/10/23
	Média	38,2857	36,2584	24,3621	29,1684	22,8222

Na maioria das vezes, o número de espectadores ao vivo tem ficado abaixo do número de matriculados, valendo lembrar que as transmissões são abertas ao público e portanto nem todos os espectadores são necessariamente discentes dos Programas de Pós-Graduação. Possíveis fatores são a troca de mediação dos seminários e divulgação interna nos respectivos Programas; do interesse dos estudantes sobre a temática do dia; da competição dessa com outras atividades acadêmicas (provas e trabalhos de outros componentes curriculares) e da possibilidade de acompanhar as transmissões em outros momentos não ao vivo, pois ficam disponíveis no canal do YouTube do ForPPGC-PR.

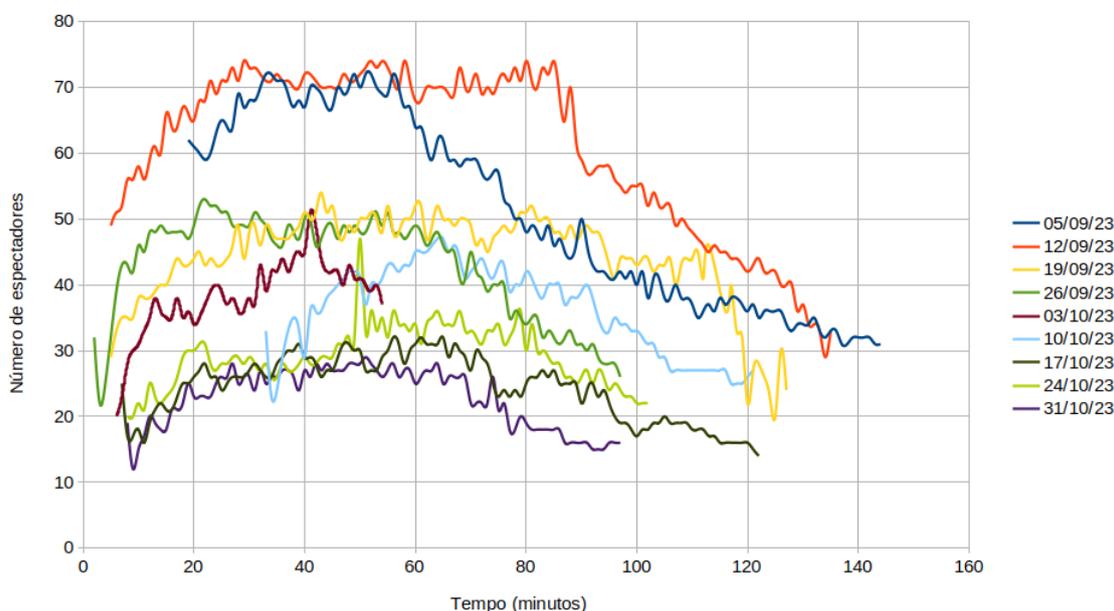


Figura 4. Número de espectadores simultâneos ao vivo ao longo do tempo das transmissões monitoradas do ForPPGC-PR em 2023-2

4.1. Questões de pesquisa levantadas em aberto

Das questões de pesquisa levantadas inicialmente, duas delas continuam em aberto: “qual a relação entre o número de visualizações ao vivo e o total de visualizações dos vídeos?” e “o número de visualizações ao vivo reflete a lista de presença assinada?”. Além dessas,

outras podem surgir que podem se relacionar às perguntas originais ou ajudar a explicá-las, como “o comportamento dos espectadores muda após o fim da apresentação (logo antes da seção de perguntas)?” e “qual o comportamento dos espectadores após o compartilhamento do *link* da lista de presenças ser compartilhado?”.

A primeira, possui um indicativo de que a quantidade de espectadores simultâneos é parecida com o total de visualizações do vídeo logo após o fim da transmissão. Ao fim da transmissão realizada em 31/10, carregou-se o *link* do vídeo para se verificar a quantidade de visualizações computadas pelo YouTube, sendo atribuída 40 visualizações.

Como o monitoramento demonstrou um mínimo e máximo de 12 e 29 espectadores simultâneos, respectivamente, além de média 22,92; desvio padrão 4,47; mediana 25 e moda 26 (16 repetições); pode-se ver que o grupo de visualizadores variou relativamente pouco em números absolutos entre o máximo de espectadores ao vivo (garantidamente espectadores únicos) e o máximo registrado pela mídia social até aquele momento.

A diferença provavelmente se caracteriza pelas entradas e saídas por motivos diversos, como perda de conexão, desinteresse pelo tema, motivos pessoais, assinatura da lista de presença, horário avançado, proximidade do fim das perguntas do *chat*. Contudo, a última pergunta em aberto original, que ainda não pode ser conferida, pode ter relação com essa variação. Para respondê-la, será necessário entrar em contato com a organização dos seminários, o que foge do escopo do monitoramento independente deste trabalho proposto.

5. Conclusão e trabalhos futuros

Nos dias atuais, a Revolução da Internet tem trazido para os ambientes virtuais muitas organizações da humanidade. Uma delas é o conceito de redes sociais, no ciberespaço chamadas de mídias sociais. Com objetivos como armazenar e trocar experiências ou gerenciar conhecimento, tem-se estudado e se analisado métricas para conferir se tais objetivos têm sido alcançados.

Os diferentes blogs (de texto, imagens ou vídeos) fazem parte de uma categoria de mídia social. Ao mesmo tempo que são meios de comunicação em massa (capaz de atingir muitas pessoas), também permite colaboração e interação de pessoas sobre uma postagem, tornando esses sistemas colaborativos.

Dos blogs, o YouTube é uma ferramenta que possui diversas métricas disponíveis apenas aos administradores do canal porém demais usuários a princípio não têm acesso a todas essas métricas. Uma delas, a popularidade dos vídeos, importante para essa categoria, não é uma possibilidade aos usuários comuns em vídeos ao vivo.

Para permitir aos usuários comuns acessarem a estatística de números de espectadores simultâneos ao longo do tempo em vídeos ao vivo, este trabalho propôs um sistema simples de monitoramento que coleta tal quantidade a cada minuto e a guarda em um banco de dados. A partir desse, é possível se construir gráficos e se calcular estatísticas sobre qualquer vídeo monitorado no YouTube mesmo sem ser administrador do canal.

Como prova de execução, o sistema construído foi aplicado aos vídeos das transmissões ao vivo do Fórum dos Programas de Pós-Graduação em Computação do Paraná (ForPPGC-PR) de setembro e outubro de 2023. Notou-se que a audiência começa a diminuir após aproximadamente 60 minutos do início da transmissão em aproximadamente

metade das transmissões, além de que o número médio de espectadores têm diminuído com o passar do semestre.

Perguntas em aberto desta pesquisa dizem respeito a acompanhar se o número de visualizações contabilizadas pelo YouTube ao fim da transmissão é próximo do número máximo registrado pelo sistema de monitoramento e qual o comportamento do espectador logo após o fim da apresentação do palestrante (antes da sessão de perguntas) e após o receber o *link* do formulário de registro de presenças.

Outros trabalhos futuros podem envolver acompanhar a evolução do crescimento do número de visualizações dos vídeos ao vivo dos canais após certos períodos de tempo depois do fim da transmissão, como imediatamente após o fim, uma semana após, um mês, seis meses e um ano.

6. Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001 e Programa de Excelência Acadêmica (PROEX). Agradece-se ao DInf da UFPR pela vivência em seu ambiente acadêmico e bolsa de doutorado. Agradece-se também ao IFC e ao IFC *campus* Brusque por oportunizar o afastamento do servidor para capacitação.

Referências

- de Lemos MEIRA, S. R., COSTA, R. A., JUCÁ, P. M., and da SILVA, E. M. (2011). Redes sociais. In Pimentel, M. and Fuks, H., editors, *Sistemas Colaborativos*. Elsevier Campus.
- NICOLACI-DA-COSTA, A. M. and PIMENTEL, M. (2011). Sistemas colaborativos para uma nova sociedade e um novo ser humano. In Pimentel, M. and Fuks, H., editors, *Sistemas Colaborativos*. Elsevier Campus.
- PIMENTEL, M., GEROSA, M. A., and FUKS, H. (2011). Sistemas de comunicação para colaboração. In Pimentel, M. and Fuks, H., editors, *Sistemas Colaborativos*. Elsevier Campus.
- Souza, G. (2023). Qual a rede social mais usada em 2023? a resposta vai te surpreender. Disponível em: <https://www.techtudo.com.br/listas/2023/07/qual-a-rede-social-mais-usada-em-2023-a-resposta-vai-te-surpreender-edapps.ghtml>. Acesso em: 06 de nov. de 2023.